ABSTRACT

MOBILITY MODELING AND MANAGEMENT FOR NEXT GENERATION WIRELESS NETWORKS

by
Woo-Jin Choi

Mobility modeling and management in wireless networks are the set of tasks performed in order to model motion patterns, predict trajectories, get information on mobiles' whereabouts and to make use of this information in handoff, routing, location management, resource allocation and other functions.

In the literature, the speed of mobile is often and misleadingly referred to as the level of mobility, such as "high" or "low" mobility. This dissertation presents an information theoretic approach to mobility modeling and management, in which mobility is considered as a measure of uncertainty in mobile's trajectory, that is, the mobility is low if the trajectory of a mobile is highly predictable even if the mobile is moving with high speed. On the other hand, the mobility is high if the trajectory of the mobile is highly erratic. Based on this mobility modeling concept, we classify mobiles into predictable and non-predictable mobility classes and optimize network operations for each mobility class. The dynamic mobility classification technique is applied to various mobility related issues of the next generation wireless networks such as location management, location-based services, and energy efficient routing in multihop cellular networks.

MOBILITY MODELING AND MANAGEMENT FOR NEXT GENERATION WIRELESS NETWORKS

by
Woo-Jin Choi

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Electrical Engineering

Department of Electrical and Computer Engineering, NJIT

August 2003

APPROVAL PAGE

MOBILITY MODELING AND MANAGEMENT FOR NEXT GENERATION WIRELESS NETWORKS

Woo-Jin Choi

Dr. Sirin Tekinay, Dissertation Advisor — Date
Assistant Professor of Electrical and Computer Engineering, NJIT

Dr. Christopher Rose, Committee Member — Date
Professor of Electrical and Computer Engineering, Rutgers University

Dr. Nirwan Ansari, Committee Member — Date
Professor of Electrical and Computer Engineering, NJIT

Dr. Constantine Manikopoulos, Committee Member — Date
Associate Professor of Electrical and Computer Engineering, NJIT

Dr. Roberto Rojas-Cessa, Committee Member — Date
Assistant Professor of Electrical and Computer Engineering, NJIT

BIOGRAPHICAL SKETCH

Author: Woo-Jin Choi

Degree: Doctor of Philosophy

Date: August 2003

Undergraduate and Graduate Education:

- Doctor of Philosophy in Electrical Engineering,
  New Jersey Institute of Technology, Newark, NJ, USA, 2003

- Master of Science in Electrical Engineering,
  Columbia University, New York, NY, USA, 1992

- Bachelor of Science in Electrical Engineering,
  Fairleigh Dickinson University, Teaneck, NJ, USA, 1990

Major: Electrical Engineering

Presentations and Publications:

Woo-Jin Choi and Sirin Tekinay,
"Predictability-Based Adaptive Location Update Scheme Using Dynamic Mobility Classification,"
submitted to *ACM Wireless Networks.*

Woo-Jin Choi and Sirin Tekinay,
"Location Based Service (LBS) Provisioning for Next Generation Wireless Networks,"
Kluwer International Journal of Wireless Information Networks, vol. 10, no. 3, 2003.

Woo-Jin Choi and Sirin Tekinay,
"Mobility-Based Energy Efficient Multihop Routing for Cellular Ad Hoc Augmented Networks,"
submitted to *Journal of Communications and Networks.*

Woo-Jin Choi and Sirin Tekinay,
"Location Based Services for Next Generation Wireless Mobile Networks,"
in *Proc. IEEE Vehicular Technology Conference Spring*, 2003.

Woo-Jin Choi and Sirin Tekinay,
"Link State Predictive Routing in Wireless Ad Hoc Networks,"
in *Proc. CISS'02, the 36th Annual Conference on Information Sciences and Systems*, Princeton University, NJ, USA, 2002, pp. 149-153.

Woo-Jin Choi and Sirin Tekinay,
"An Adaptive Location Registration Scheme with Dynamic Mobility Classification,"
in *Proc. IEEE International Conference on Communications*, 2002, pp. 440-444.

Woo-Jin Choi and Sirin Tekinay,
"Distance-Based Location Update Scheme with Isotropic Random User Motion with Drift,"
in *Proc. IEEE Vehicular Technology Conference Spring*, 2002, pp. 1111-1115.

Woo-Jin Choi and Sirin Tekinay,
"An Efficient Table Driven Routing Algorithm for Wireless Ad Hoc Networks,"
in *Proc. IEEE Vehicular Technology Conference Fall*, 2001, pp. 2604-2608.

Woo-Jin Choi and Sirin Tekinay,
"Mobility Modeling and Management for Next Generation Wireless Networks,"
in *Proc. ISCIS'01, the 16th International Symposium on Computer and Information Sciences*, Antalya, Turkey, 2001, pp. 517-525.

Woo-Jin Choi and Sirin Tekinay,
"Mobility Modeling and Management for Next Generation Wireless Networks,"
in *Proc. WPMC'01, the 4th International Symposium on Wireless Personal Multimedia Communications*, Aalborg, Denmark, 2001.

I am blessed with a wonderful family
to whom this work is dedicated

To my wife Jung-Hyun Kang
and our child Jhee-Woo Choi

ACKNOWLEDGMENT

I would like to express my deepest appreciation to Dr. Sirin Tekinay at NJIT, who not only served as my research advisor, providing valuable and countless resources, and insight, but also constantly gave me support and encouragement. Special thanks to Dr. Christopher Rose at Rutgers University, Dr. Nirwan Ansari, Dr. Constantine Manikopoulos, Dr. Roberto Rojas-Cessa and Dr. Symeon Papavassiliou at NJIT for actively participating in my dissertation committee providing valuable comments and guides.

My fellow students at New Jersey Center for Wireless Telecommunications and New Jersey Center for Wireless Networking and Internet Security deserve the recognition for their support.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1
INTRODUCTION

Mobility modeling in wireless networks is to estimate the impact of mobility on the wireless systems and to evaluate the performance of resource allocation, handoff, admission control and location management techniques. Collecting, processing and making use of the mobility and location related information of mobile users are tasks usually performed by central or distributed agents operating the network. Optimization on the algorithms keeping track of users' location for the purpose of delivering applications or services is the main theme of mobility management. The specific purpose of analysis or optimization dictates the approach to be taken and the complexity of the mobility model. This dissertation introduces a new mobility modeling approach: Predictability-based Mobility Modeling. This new mobility concept is used to optimize mobility management of the next generation wireless networks. This chapter provides an overview of the conventional mobility modeling techniques and presents a vision of the next generation wireless networks that call for a new breed of mobility modeling and management techniques.

1.1 Conventional Mobility Modeling Techniques

In the conventional wireless networks, the primary application was voice. Depending on the scope of mobility in a specific network operation that delivers calls or maintains on-going call, the mobility modeling techniques are designed to evaluate the performance of the operation. For example, the power allocation to combat intra-cell interference in spread spectrum system concerns the location of users in a single cell. The resource reservation for handoff users concerns the mobility of users in a group of neighboring cells [1]. Per-user location management schemes concern the mobility of individual users. If there is a mobility model that concerns with individual as well as collective characteristics of users' motion pattern, the model should describe the movement of each user and the spatial and temporal correlation among different users. To cope with the obvious scalability problem, researchers limit their scope of analysis to specific network operation. Simplifying assumptions like stationary mobility (i.e., statistical properties that describe user mobility doesn't change in time), uniform mobility (i.e., the same mobility characteristics are assumed for all users) and uniform user population (i.e., the density of mobile users is uniform over the entire network) are often introduced.

1.1.1 Constant Speed/Direction Mobility Model

As long as the movement of individual user is concerned, the simplest mobility model is the constant speed and constant direction model in [2]. The speed of a mobile is assumed to be constant, and only four orthogonal directions of motion along the Cartesian axes are allowed. A mobile is equally likely to start moving in any of these four directions at call initiation. The mobile keeps the same speed and direction for the entire duration of a call. The handoff probability, which is the probability that a call goes through at least n handoffs during a call, is given as a function of the cell radius, the speed of mobile, and the average call duration (See equation 4 and 5 in [2]). Since the channel occupancy time is smaller than both the call duration and the time that the call is handed off to another cell, the channel occupancy time distribution is calculated based on these two events that are not mutually exclusive. It is shown in [2] that the channel occupancy time distribution is very close to an exponential distribution.

1.1.2 Random Waypoint Mobility Model

In the random waypoint mobility model [3], each mobile selects a random destination in a given area and moves to the destination at a speed which is drawn from a uniform distribution between 0 and some maximum speed. Arriving at its destination, the mobile stays put for a time duration called pause time. When the pause time is over, the mobile selects another destination and speed, and starts moving again.

1.1.3 Random Walk Mobility Model

In the random walk mobility model [4], it is assumed that individual user's cell transition probability from one cell to every other cell is known. A concept called shadow cluster is introduced where the shadow is comprised of individual user's resource requirements along the predicted path of the user. Time-dependent cell transition probability is used to calculate the resource requirement of each user. Admission control is performed based on the current local traffic condition and the predicted resource requirement for mobiles in the local area.

The mobility model in [5] uses time-invariant cell to cell transition probability which is assumed to be known to the network. A mobile's dwell time in a cell is exponentially distributed with parameter that depends on the user and the cell. The future cell occupancy probability of a user is predicted using a given cell transition probability matrix, and the location area of a user is selected based on the predicted cell occupancy probability.

The resource allocation and admission control in [6] requires each mobile to provide mobility information to the network. The mobility information consists of a set of cells that the user is expected to visit during the life time of its connection.

In the symmetric random walk model, time is slotted with equal length, and mobile makes at most one move during each time slot. In the one-dimensional version of symmetric random walk, a mobile moves to a neighboring cell in the right with probability p and moves to a neighboring cell in the left with the same probability p and stays in the current cell with probability $1-2p$. In two-dimensional random walk with mesh topology, each cell has 4 neighboring cells. The probability that a mobile stays put in the current cell during the next time slot is $1 - 4p$. In the hexagonal topology where each cell has 6 neighboring cells, the probability that a mobile stays in the current cell is $1 - 6p$. The symmetric random walk model has been used extensively since it simplifies the analysis; however, a drawback of the symmetric random walk model is that the movement of mobile is not biased in direction.

1.1.4 Brownian Motion Model

Brownian mobility model is used to describe movements of an individual user in [7] [8]. In the one-dimensional version, a mobile user moves one "space step" Δx to the right with probability p and to the left with probability q, and stays at the current position with probability $1-p-q$ for each "time step" Δt. Assuming that a mobile started at time $t = 0$ at position $x = 0$, as the time and the space steps are made infinitesimally small the displacement of the mobile at time $t(> 0)$ is described by a Gaussian density function:

$$p_{X(t)}(x(t)) = \frac{1}{\sqrt{\pi Dt}} \cdot e^{-(x-\mu t)^2/Dt} \tag{1.1}$$

where $\mu = (p-q)\Delta x/\Delta t$ is the drift velocity and $D = 2((1-p)p + (1-q)q + 2pq)(\Delta x)^2/\Delta t$ is the diffusion constant of the Brownian motion, both functions of the time and the space steps [9]. For n-dimensional Brownian motion, the movement of an individual mobile user is described by

$$p_{X(t)}(\mathbf{x}(t)) = \frac{1}{(\pi Dt)^{n/2}} \cdot e^{-|\mathbf{x}-\mu t|^2/Dt}. \tag{1.2}$$

In the Brownian mobility model, the variance of the Gaussian displacement distribution, which is a measure of location uncertainty, is independent from the mean drift velocity μ. In other words, the location uncertainty of high speed mobiles are not necessarily large. In the dynamic location update scheme [10], mobile users are classified based on the uncertainty in their motion pattern, and each mobile is assigned a different location update scheme, that is, a predictive scheme for mobiles whose trajectory is highly predictable and a non-predictive scheme for mobiles whose trajectory is highly erratic.

1.1.5 Fractional Brownian Motion Model

The empirical transportation engineering studies [11] show that the mobile speed at a given point in time has a Gaussian distribution where the mean is equal to the speed limit of the street. Although the conventional Brownian motion model leads to a Gaussian displacement distribution, in reality the mobile users are street bound and speed regulated by the local speed limit, that is, mobiles in one street may accelerate their speed while mobiles in another street may decelerate due to the geography of street layout. The fractional Brownian motion (FBM) mobility model in [12] is a generalization of the Brownian motion. The displacement of a mobile is described by Gaussian distribution with mean $\mu \cdot t^{\alpha}$ and variance $\sigma^2 \cdot t^{2\alpha}$ where α $(0 \leq \alpha \leq 1)$ is the Hurst parameter of the fractional Brownian motion. For $\alpha = 0.5$, FBM models non-accelerating motion patterns. For $\alpha > 0.5$, FBM describes accelerating motion pattern. For $\alpha < 0.5$, FBM describes decelerating motion patterns.

1.1.6 Fluid Flow Mobility Model

The fluid flow mobility model [13] describes macroscopic movement behavior by modeling traffic flow as if it is a flow of fluid. The amount of mobile traffic moving out of a region is assumed to be proportional to (a) the average speed of mobiles, (b) the length of the region boundary, and (c) the density of mobiles in the region. Assuming that the movement direction of mobiles is uniformly distributed over (0,

2π), the average number of mobiles moving out of a region per unit time is described as

$$n = \frac{\rho \cdot v \cdot \ell}{\pi} \tag{1.3}$$

where

ρ is the density of mobile (mobiles/unit area),

ℓ is the length of the region boundary,

v is the average speed of mobiles.

In the dynamic location management scheme [14], the fluid mobility model is used to determine the number of location updates per mobile per unit time for different location area size. Dividing the number of mobiles moving out of a location area per unit time by the average number of mobiles in the location area, the number of location updates per mobile per unit time becomes

$$f = \frac{n}{\rho A} = \frac{v\ell/\pi}{A} \tag{1.4}$$

where ℓ is the length of location area boundary and A is the size of location area. Assuming square shaped location area (i.e., $A = \ell^2/16$), the location update frequency per mobile becomes

$$f = \frac{v/\pi}{\ell/16}. \tag{1.5}$$

When the size of the location area is large (i.e., ℓ is large), the location update frequency (1.5) is small, but the size of the paging area, in which the network pages a mobile upon a call arrival, becomes large. If the size of the location area is small (i.e., ℓ is small), the paging area becomes small, but the location update frequency becomes large. The optimal location area size that minimizes overall signaling cost is determined in [14].

In the location area optimization problem [15], the same fluid mobility model is used. The shape of cell is assumed to be circular, and the rate of cell boundary crossing is described as

$$f_{cell} = \frac{2v}{\pi r} \tag{1.6}$$

where r is the radius of a cell. By scaling the cell boundary crossing rate by $P_{LA/cell}$ which is the proportion factor of the perimeter of a cell that coincides with the parameter of a location area (LA), the number of location updates per unit time is described as

$$f_{LA} = \frac{2v}{\pi r} \cdot P_{LA/cell} \tag{1.7}$$

1.1.7 Gravitational Mobility Model

The gravitational mobility model conceptualizes mobile traffic as the gravity between two objects in Physics. The amount of mobile traffic moving from one region to another is assumed to be proportional to the population in both the region that mobiles are moving out of, and the region that the mobiles are moving into. The gravitational model was developed for transportation research to model human movement behavior. Assuming that N_i is the population in region i, and $K_{i,j}$ is a "gravitational" parameter for region i and j, which has to be determined from field data, the amount of traffic moving from region i to region j is described as

$$T_{i,j} = K_{i,j} \cdot N_i \cdot N_j \tag{1.8}$$

A variation of this model is used for performance analysis for PCS networks in [16]. The mobility model is used to characterize movements of mobiles on different scales: within a metropolitan area, within a national area, and at the international level. The performance analysis on IS-41, the current location management standard in the United States, is carried out for San Francisco Bay area.

1.1.8 State Space Mobility Model

In the state space mobility model introduced in [17], the handoff probability is determined by modeling user mobility using multi-dimensional Markov chain where each state consists of the number of mobiles in each cell. The dwell time of a mobile in a cell is assumed to be exponentially distributed with parameter that is dependent on the cell. The number of states that describe system operation increases exponentially as the number of cells in the network increases:

$$s = (1 + c)^n \tag{1.9}$$

where

s is the number of states,

c is the maximum number of available channels per cell assuming Fixed Channel Allocation (FCA) [18],

n is the number of cells in the network.

A drawback of this approach is that the number of states increases exponentially as the number of cells in the network increases. This problem is noted as state space explosion in [19].

In Hidden Semi-Markov mobility model [20], the state of mobile is defined in terms of a vector, and each component of the vector represents a value from finite attribute space S where the attribute space represents the properties of mobiles such as location, direction, speed, etc. The "motion" of mobile is described by its trajectory in the attribute space. The state transition of a mobile is described by Markov chain. The aggregate behavior of mobiles is represented by a vector process

$$N(t) = (N_1(t), ..., N_m(t), ..., N_M(t)) \tag{1.10}$$

where $N_m(t)$ is the number of mobile users in state m at time t. The steady-state probability distribution of the state vector is given as a product form solution.

1.1.9 Cell Residence Time Distribution

The cell residence time and the channel holding time distribution of a user is often assumed to be exponential. The validity of the assumption is shown in [2] based on computer simulation. In the simulation, each cell is assumed to be identical and circular in shape, and the speed of mobile is assumed to be constant. At call initiation, a mobile is assigned (a) a random initial position in a cell drawn from a uniform distribution, (b) a random initial direction drawn from a uniform distribution over $(0,2\pi)$ and (c) a random time until the direction changes. The random time between the changes in the direction has an exponential distribution assuming that the changes occur at memoryless instants. The lack of memory in motion pattern suggests exponential distribution of the cell residence time. The call holding time is also assumed to be exponentially distributed. It is interesting to note that the average rate of changes in direction should be kept in a certain range to preserve the memoryless property. It is shown that if the rate is either too small or too large, the channel holding time distribution deviates from the exponential distribution. In fact, if the mobile doesn't change direction at all, the trajectory of mobile is deterministic; therefore, the cell residence time is also deterministic. If the mobile changes direction frequently, the movement of mobile approximates a Brownian motion process. An analytical expression for cell residence time distribution is derived in [21] based on a similar set of assumptions. It is shown that the cell residence time distribution strongly resembles the exponential distribution.

Call blocking probabilities are calculated assuming the exponential channel holding time and the exponential call holding time distribution in [22] [23] [24]. The sequence of cell entrance times of mobiles are modeled as a Poisson process with rate γ in [25]. The number of mobiles residing in a cell is described using M/G/∞ queuing system, and the average number of mobiles per cell is calculated using Little's theorem: $n_{MS} = \gamma \cdot E\{S\}$ where S is cell residence time. Since the average number of mobiles entering a cell per unit time is γ, the cell boundary crossing rate per mobile is calculated as

$$\begin{aligned} \lambda_{cell} &= \gamma / n_{MS} \\ &= 1/E\{S\}. \end{aligned} \tag{1.11}$$

The location area boundary crossing rate is obtained by weighting the cell boundary crossing rate by a ratio between the perimeter of location area and the perimeter of cell as in the fluid flow mobility model (i.e., the boundary crossing rate increases as the perimeter of the region increases).

The biased sampling problem of handoff mobiles is pointed out in [26]. Since high speed mobiles are more likely to cross cell boundaries than low speed mobiles, the speed of hanfoff mobiles observed at cell boundaries is higher than the speed of an average mobile, and the cell residence time of handoff mobiles is smaller than that of an average mobile. The handoff blocking probability is analyzed in [27] considering the effect of the bias on channel holding time distribution, and the accuracy of the analysis is demonstrated through simulation.

In reality, the shape of cell is irregular, and mobile's speed, direction and time between changes in direction are difficult to characterize. There has been attempt to model the cell residence time distribution based on field data. A generalized Gamma distribution is proposed to model the cell residence time in [28]. The cell residence time is modeled using SOHYP (the sum of Hyperexponential) distribution in [29]. The hyper-Erlang distribution, which is known for its capability of universal approximation, is proposed to model the cell residence time in [30].

1.2 Challenges of the Next Generation Wireless Network

In the conventional wireless networks, mobility modeling is used to characterize the behavior of users on their physical location as well as the behavior of the criteria (e.g., observed power level, signal to noise ratio) that give rise to the execution of mobility related transactions. In most cases, it is assumed that there is a one-to-one mapping between the two, which implies that the major factor governing wireless transmission is the physical distance between transceivers. Indeed, in the absence of information regarding peculiarities of the propagation environment (such as multipath scattering or the absence of line of sight signals), this is a viable assumption for purposes of network performance modeling. The boundary of interest in mobility modeling can be a geographic boundary (e.g., sector, cell and location area boundaries), or it can be a threshold of a measure that displays stochastic behavior such as the received power level where the threshold for handoff decision is given. Therefore, boundary crossing models are used to estimate the probability of, or the rate of, inter-sector, inter-cell or inter-system handoffs, and location updates.

As the service demand becomes more and more data oriented along with the advent of proximity based wireless systems, peer-to-peer communications, and ad hoc networks, the conventional voice oriented wireless service architecture limits the freedom that the wireless can bring in the future. Freedom of being mobile is not simply that the wires attached to handsets are removed, but it claims for a new dimension of wireless service in which wireless information delivery is custom-tailored to where users are and what users are doing. In other words, the geographic relevance of application becomes an important service quality. In the next generation wireless networks, the concept of user location includes not only the ID of the base station that the user is associated with, but also the geographic position and the state of mobility, that is, whether the user is in vehicle, at work, at home, etc.

An example of location-aware application service is a geographically targeted multicasting service that is called geomulticasting. Due to the locality of the target subscribers, a base station does not need to deliver multiple copies of a same message to each user in the target area. Instead, users in the vicinity may relay or share the message through peer-to-peer communications. One candidate for such network architecture is Cellular Ad Hoc Augmented Network (CAHAN) [31]. CAHAN is a hybrid network, in which ad hoc networking capability is added to mobile stations. Typically, the range for peer-to-peer radio communications is much shorter than the radius of a cell, and multiple ad hoc networks may exist in a single cell.

Another type of location-aware application service is geocasting in which a user may send a message or establish communication with mobiles or fixed entity in a target area of interest. In this application, the users don't care who they are talking to, but they really care that they are talking to someone in the area of interest. A gas station may want to advertise its special rate to vehicular users passing through its local area. Since the message is targeted to vehicular users, the message should be delivered to users who are moving at vehicular speed. A user passing through the target area should receive the advertisement message during the stay in the target area.

The concept of the location-aware application service is simple, and each component technology required to realize the service is already there: the geolocation techniques such as GPS [32] and Indoor Geolocation [33], and the wireless network infrastructure. However, the challenges in the wireless networks is to collect the geographic position and the mobility state information from mobile users and to provide this information to application services so that the services can be customized to users' need.

The predictability-based mobility modeling approach and user tracking strategies introduced in this dissertation are applicable to virtually every mobility related transactions in wireless networks. Usefulness of the new mobility modeling concept is demonstrated for location management, location-based service provisioning, and routing in multihop cellular networks.

1.3 Dissertation Outline

Chapter 2 introduces a predictability-based mobility modeling approach. In the predictability-based approach, the mobility is defined as the uncertainty in mobiles' motion pattern. A high speed mobile could have small uncertainty in its motion pattern if its trajectory is highly predictable. A slow mobile could have large uncertainty in the motion pattern if its trajectory is highly erratic. It is shown that the predictability-based mobility modeling will lower the cost of the mobility related transactions in the wireless network by predicting the trajectories of mobiles that are highly biased in direction.

Chapter 3 introduces a predictability-based location update scheme in which the predictive and the non-predictive distance-based location update schemes are dynamically selected for each mobile. A mobile with highly predictable trajectory is assigned a predictive location update scheme, and a mobile with random trajectory is assigned a non-predictive location update scheme. The analytical framework is presented for optimal mobility classification.

Chapter 4 introduces a hybrid geolocation update scheme for location-based application services (LBS) in wireless networks. The proposed geolocation update schemes reduce the geolocation update frequency and satisfy the QoS of application service. For case studies, Location-based Traffic Report Service (LBS-TR) and Location-based Navigation Service (LBS-NS) are presented. In LBS-TR, mobile users who are about to enter heavily jammed highways are informed in a timely manner so that the users can avoid traffic congestion. In LBS-NS, mobile users are provided with navigation instructions that are custom-tailored to the geolocation of the users. The granularity of navigation instruction changes as the user moves from highway to local street and from local street into building complex.

Chapter 5 presents a mobility-based network protocol that reduces energy consumption for multihop cellular networks. A mobile makes routing decision based on the movement direction of its neighboring mobiles. If source mobile S has a neighboring mobile, R, which is moving towards a cellular base station, then S relays its packets to R so that R can deliver the packets to the base station when the position of R is close to the base station. On its way to the base station, if R encounters another mobile, R', which will approach the base station even "closer" than R, R relays the packets to R'. R' will do the same when it finds a "better" relay mobile. Under reasonable mobility assumptions, the simulation result shows that the proposed mobility-based multihop routing consumes less than 0.044% of the energy required by the conventional single-hop routing for upstream traffic.

The conclusion of the dissertation is given in Chapter 6.

CHAPTER 2

PREDICTABILITY-BASED MOBILITY MODELING

2.1 Predictability as a Measure of Mobility

User mobility is one of the most salient characteristic in wireless mobile networks. Mobility causes complexities in the network operations and system architecture. Mobility modeling has played an important role in examining various issues in network operations. In the literature, the average speed of mobile is often referred to as the level of mobility, such as "high" or "low" mobility. This chapter introduces a new mobility modeling concept that defines mobility as the uncertainty of mobile's trajectory. The mobility of a high speed mobile is not necessarily high if the trajectory of the mobile can be predicted accurately. A mobile on a highly directed path such a vehicle on a highway usually travels long distance before changing its speed and direction. The network can predict the trajectory of the mobile using its speed and direction [34]. Unless the mobile changes the speed or the direction, the mobile does not need to be inform the network about the mobile's location or mobility state to receive and/or customize application services.

Trajectory prediction is useful for many issues in mobility management in wireless networks. In ad hoc networks, each mobile station functions as a router, and the network topology is heavily dependent on the location of each mobile. With a cost effective geolocation information dissemination scheme, each mobile can determine multihop routes to other mobiles in a fully distributed manner. In the mobility-based location information dissemination scheme [35], each mobile disseminates messages containing its current location information at a rate that is optimized according to the speed of each mobile, i.e., the rate of location update is small when the speed of the mobile is low, and the rate becomes large when the speed becomes high. With the trajectory prediction, a high speed mobile does not necessarily update its location frequently as long as the mobile maintains its speed and direction.

In the location management for cellular networks, each mobile can register its location and velocity (i.e., speed and direction) during the location registration process [36]. Since the network can predict the location of a mobile based on the speed and direction information, the mobile may register its location again when the distance between the current location and the predicted location exceeds a predetermined threshold.

A predictive QoS provisioning for wireless cellular network is proposed in [37]. The network predicts the direction of each mobile as well as the time of arrival and departure at those cells that the mobile is likely to visit in the future. The network reserves radio channels in those cells just for the duration between the predicted arrival time and departure time. Therefore, a single radio channel can be reserved for a number of different mobiles as long as the mobiles are not scheduled to be at the same cell at the same time.

Despite the complexity in trajectory prediction, the resource allocation based on per user mobility becomes more attractive as the wireless evolution continues. Since the future wireless network is expected to provide more wide-band services to a greater number of mobile users, the radio communication range will become smaller and smaller for the efficient reuse of the limited bandwidth spectrum. Since the number of radio connections that each base station can provide to mobile users will be small, the coarse granularity of radio resource becomes difficult to manage without considering individual user mobility. It is obvious that if the motion pattern of a mobile is highly predictable, then the resource allocation based on the trajectory prediction will lower the handoff dropping probabilities.

The trajectory prediction will lower the network operation cost if the motion pattern of mobiles is highly predictable. If the motion pattern becomes erratic, then one can no longer expect to benefit from the trajectory prediction. For the following analysis, the Brownian motion mobility model is used to determine the feasibility of the trajectory prediction.

2.2 Brownian Motion Model

Brownian motion process is a continuous time, continuous space, Markov process. The Brownian motion process models individual user mobility. The Brownian motion model is superior to other mobility models in the following sense:

- Continuous State Space - The movement is not restricted to Cartesian axes as opposed to the symmetric random walk model [38][39], the discrete-time Brownian motion model [7][8] and Markov chain models [40].

- Drift Movement - The direction of mobile's movement on a highly directed path can be described as continuous-time Markov process.

2.2.1 Brownian Motion without Drift

Brownian motion without drift is a zero mean diffusion process $\{X(t); t \geq 0\}$ that has following properties [41]:

(a) Every increment $X(t+s) - X(s)$ has normal distribution with mean zero and variance $\sigma^2 t$ where σ is constant independent of time t.

(b) For every pair of disjoint time intervals (t_1,t_2), (t_3,t_4), where $t_1 > t_2 \geq t_3 > t_4$, the increment $X(t_4) - X(t_3)$ and $X(t_2) - X(t_1)$ are independent random variables with distribution given in (a), and similarly for n disjoint time intervals where n is an arbitrary positive integer.

(c) $X(t)$ is continuous at t=0 where $X(0) = 0$.

2.2.2 Modified Brownian Motion with Drift

Let $\{\widetilde{X}(t), t \geq 0\}$ be a Brownian motion without drift. Brownian motion with drift is a stochastic process having the distribution of

$$X(t) = \widetilde{X}(t) + \mu t, \quad t \geq 0, \tag{2.1}$$

where μ is a constant called the drift velocity. Brownian motion with drift is a stochastic process $\{X(t); t \geq 0\}$ with the following properties [41]:

(a) Every increment $X(t+s) - X(s)$ has normal distribution with mean μt and variance $\sigma^2 t$ where μ and σ are constants independent of time t.

(a) For every pair of disjoint time intervals (t_1,t_2), (t_3,t_4), where $t_1 > t_2 \geq t_3 > t_4$, the increment $X(t_4) - X(t_3)$ and $X(t_2) - X(t_1)$ are independent random variables with distribution given in (a), and similarly for n disjoint time intervals where n is an arbitrary positive integer.

(a) $X(t)$ is continuous at t=0 where $X(0) = 0$.

The Brownian motion process is denoted as B(μ, σ) where μ is the drift velocity and σ^2 is the variance parameter. The Brownian motion do not have drift when $\mu = 0$.

2.3 Feasibility Region of the Trajectory Prediction

In this section, the feasibility of the trajectory prediction is investigated. Comparing the performance of a predictive and a non-predictive user tracking strategy, the feasibility region of the trajectory prediction is determined. Two user tracking strategies are (a) Non-Predictive Distance-based Geolocation Update Scheme and (b) Predictive Distance-based Geolocation Update Scheme. Under the assumption that one cannot have perfect knowledge on the statistics of users motion pattern, that is, the trajectory prediction is based on users speed and direction where the speed and direction have to be measured in reality, the predictive scheme does not always perform better than the non-predictive scheme in terms of the average number of geolocation updates per unit time. Under certain mobility assumptions, the non-predictive scheme performs better. These two schemes are analyzed, and an optimal way to choose appropriate scheme for each mobile is proposed.

2.3.1 Non-Predictive Distance-based Geolocation Update Scheme

When mobiles are highly biased in direction, the future geolocation of the mobiles can be predicted using their speed and direction. In fact, if a mobile does not change its speed and direction, then the mobile does not need to update its geolocation. The network can predict the geolocation of the mobile using the speed and direction information. On the other hand, if the mobile changes its speed and/or direction frequently, then the location prediction based on the short term history of mobile's movement may not work well in terms of the prediction accuracy. If the mobile reports its speed and direction to the network whenever there is a change in its speed or direction, the network may be able to predict the geolocation of the mobile accurately; however, this requires frequent speed or direction updates from the mobile. Therefore, in the non-predictive distance-based geolocation update scheme (NDB), mobiles report their geolocation (i.e., position coordinates) [1] to the network when the distance between the current position and the last reported position becomes a predetermined threshold r as in Figure 2.1. The distance threshold r can be considered as the maximum position error (i.e., the maximum difference between the mobiles actual position and the position which is known to the network).

[1]It is assumed that the position coordinates are available using GPS [32] or other geolocation techniques [42].

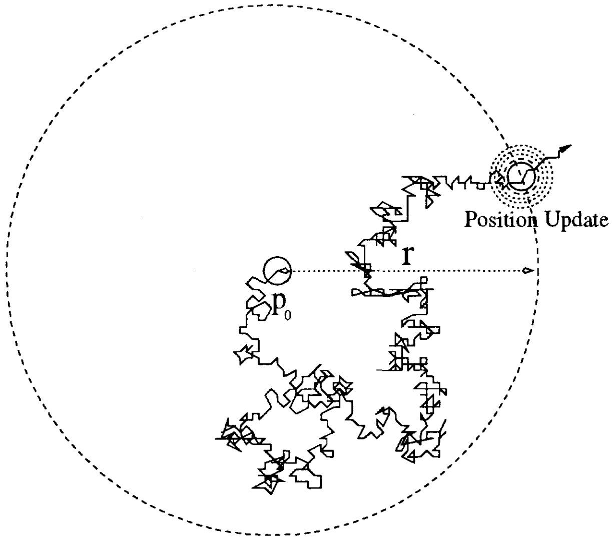


Figure 2.1 Non-predictive distance-based geolocation update scheme.

Let M denote the time between geolocation updates of a mobile, and assume that the movement of the mobile is described by two-dimensional Brownian motion $\mathrm{B}(\mu, \sigma)$. The initial position of the mobile is (0,0) at time $t = 0$, and the mobiles geolocation is assumed to be updated at time $t = 0$. Since the mobile will update its geolocation again when the distance between its position and the initial position becomes r, the time between geolocation update is equal to the dwell time of the mobile in the circular region of radius r centered at the initial position. The dwell time of the mobile is referred to as "hitting time" [43] of the Brownian motion. The Laplace-Gegenbauer transform of the hitting time distribution in n-dimensional Brownian motion is derived in [44]. In the derivation, the author assumes that the variance parameter of the Brownian motion is unity (i.e., σ^2=1). By removing this assumption, the two-dimensional version of the Laplace-Stieltjes transform of the hitting time distribution can be derived [2] as

$$\phi_M(s) = \frac{I_0(|\mu| r/\sigma^2)}{I_0(\sqrt{2s + |\mu|^2/\sigma^2} r/\sigma)} \tag{2.2}$$

where

[2] $\phi_M(s, \mu, \sigma, r) = \phi_M(s, \mu/\sigma, 1, r/\sigma)$.

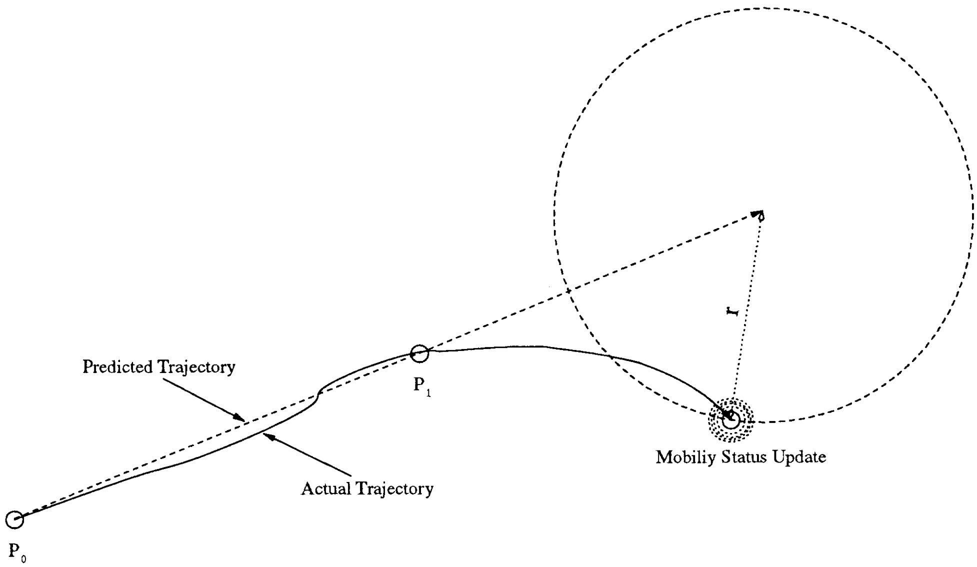


Figure 2.2 Predictive distance-based geolocation update scheme.

- M is the hitting time of Brownian motion $\mathrm{B}(\mu, \sigma)$ with circular boundary of radius r.

- r is the maximum position error.

- I_n is modified Bessel function of the first kind and order n.

Using the moment theorem, the average time between geolocation update becomes

$$\begin{aligned} E\{M\} &= \left[-\frac{d}{ds}\, \phi_{\,M}(s) \right]_{s=0} \\ &= \frac{r \cdot I_1(|\mu| r/\sigma^2)}{|\mu| \cdot I_0(|\mu| r/\sigma^2)}. \end{aligned} \tag{2.3}$$

Defining the geolocation update frequency as the inverse of the average time between two consecutive geolocation updates, the geolocation update frequency of the non-predictive scheme (NDB) becomes

$$f_{NDB}(r, |\mu|, \sigma) = \frac{|\mu| \cdot I_0(|\mu| r/\sigma^2)}{r \cdot I_1(|\mu| r/\sigma^2)}. \tag{2.4}$$

2.3.2 Predictive Distance-based Geolocation Update Scheme

Mobile with highly predictable trajectory may report its mobility status (i.e., position, speed and direction) at the time of geolocation update. The network will predict the future position of the mobile using this information. Since the position prediction is solely based on the last reported position coordinates and velocity, the mobile is aware of its predicted position at all times. The mobile may update its geolocation when the distance to the predicted position becomes a predetermined threshold r as in Figure 2.2. The distance threshold r is the maximum difference between the mobiles current position and the position which is known to the network (i.e., the predicted position).

Consider a mobile with two-dimensional Brownian motion pattern $\mathrm{B}(\mu, \sigma)$. Assume that the mobile is located at position (0,0) at time $t = 0$, and that the mobile reported its position coordinates and velocity V_0 at time $t = 0$. The mobile is assumed to determine its velocity by monitoring its displacement during the last T units of time before the geolocation update. By letting $\{Z(t), t \geq 0\}$ be a Brownian motion process without drift, $\mathrm{B}(0, \sigma)$, the position of the mobile can be expressed as a stochastic process as in (2.1):

$$X(t) = Z(t) + \mu t, \quad t > 0. \tag{2.5}$$

Since the predicted position of the mobile is

$$Y(t) = V_0 t, \quad t > 0, \tag{2.6}$$

the relative motion of the mobile with respect to the predicted position becomes

$$X(t) - Y(t) = Z(t) + (\mu - V_0)t, \quad t > 0, \tag{2.7}$$

which is another Brownian motion process having drift velocity $\mu - V_0$ and variance parameter σ^2. Since the mobile updates its geolocation when the distance to its predicted position becomes r, the time between geolocation update is equal to the hitting time of Brownian motion $B(\mu - V_0, \sigma)$ with a circular boundary of radius r. The author refers to $\mu - V_0$ as velocity jitter at time $t = 0$ and denotes the velocity jitter as

$$\nu = \mu - V_0. \tag{2.8}$$

With a given velocity jitter, the average time between geolocation update in the predictive distance-based scheme is

$$E\{M|\nu\} = \frac{r \cdot I_1(|\nu|r/\sigma^2)}{|\nu| \cdot I_0(|\nu|r/\sigma^2)}. \tag{2.9}$$

Since the displacement of the mobile during the velocity monitoring interval, which is (i.e., $V_0 T$), has a Gaussian distribution with mean μT and variance $\sigma^2 T$, νT $(= \mu T - V_0 T)$ has Gaussian distribution with mean 0 and variance $\sigma^2 T$. Therefore, the velocity jitter ν has Gaussian distribution with mean 0 and variance σ^2/T, and the amplitude of the velocity jitter, $|\nu|$, has a Rayleigh distribution [45],

$$p_Y(y, \sigma, T) = \frac{y}{\sigma^2/T} \cdot \mathrm{e}^{-\frac{y^2}{2\sigma^2/T}}, \tag{2.10}$$

where $Y = |\nu|$. Therefore, the average time between geolocation updates becomes

$$E\{M\} = \int_0^\infty \frac{r \cdot I_1(yr/\sigma^2)}{y \cdot I_0(yr/\sigma^2)} \cdot p_Y(y, \sigma, T)dy \tag{2.11}$$

where

- r is the maximum position error,
- σ^2 is the variance parameter of the Brownian motion,
- T is the velocity monitoring interval,
- I_n is modified Bessel function of the first kind and order n.

The average geolocation update frequency of the predictive scheme (PDB) becomes

$$f_{PDB}(r, \sigma, T) = \left[\int_0^\infty \frac{r \cdot I_1(yr/\sigma^2)}{y \cdot I_0(yr/\sigma^2)} \cdot p_Y(y, \sigma, T)dy\right]^{-1} \tag{2.12}$$

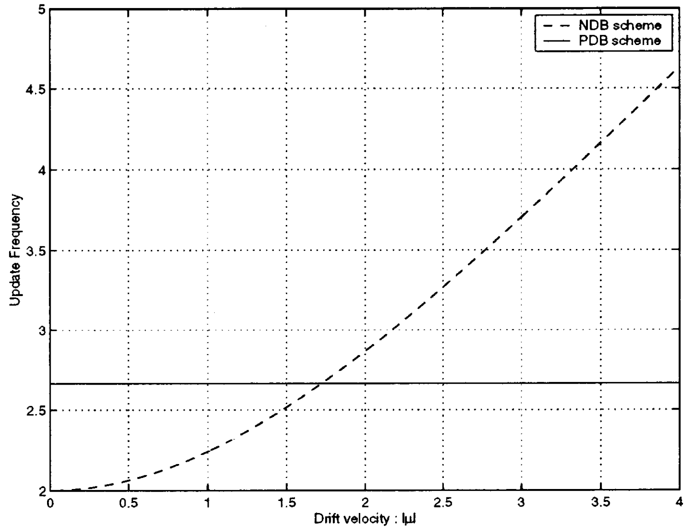


Figure 2.3 Geolocation update frequency.

2.3.3 Dynamic Mobility Classification: An Optimal Way to Minimize Geolocation Update Frequency

In the previous subsections, two geolocation update schemes are introduced and analyzed. The predictive distance-based geolocation update scheme is prescribed for mobiles with predictable trajectory, and the non-predictive distance-based geolocation update scheme is prescribed for mobiles with random trajectory. However, how can one decide whether a mobile's trajectory is predictable or unpredictable? Instead of measuring the predictability of trajectory, the predictability of a mobile's trajectory is defined in the following mobility classification rule that minimize geolocation update frequency.

A mobile's trajectory is *predictable* if $f_{NDB}(r, |\mu|, \sigma) > f_{PDB}(r, \sigma, T)$.

Otherwise, the mobile's trajectory is *unpredictable.*

In Figure 2.3, the geolocation update frequencies of the predictive and the non-predictive schemes are plotted as a function of drift velocity. The variance parameter is assumed to be $\sigma^2 = 1$, and the velocity monitoring interval is $T = 0.5$. The maximum geolocation error, which is the upper bound on the difference between a mobile's current geolocation and the nominal geolocation known to the network or predicted by the network, is $r = 1$. The dotted curve indicates the geolocation update frequency for the non-predictive scheme. The update frequency of the non-predictive scheme increases as the drift velocity increases. The solid line indicates the geolocation update frequency of the predictive scheme. The update frequency of the predictive scheme is constant because the geolocation update frequency in (2.12) is independent of drift velocity. With the mobility classification, a mobile may dynamically choose appropriate geolocation update scheme as its motion pattern changes. When $f_{NDB}(r, |\mu|, \sigma)$ is smaller than $f_{PDB}(r, \sigma, T)$, a mobile should choose the non-predictive scheme. When $f_{PDB}(r, \sigma, T)$ is smaller than $f_{NDB}(r, |\mu|, \sigma)$, the mobile should choose the predictive scheme to update its geolocation.

CHAPTER 3

PREDICTABILITY-BASED LOCATION MANAGEMENT IN CELLULAR NETWORKS

3.1 Introduction

The cellular network is the primary choice for the next generation wireless network architecture due to the wide area coverage. One of the important issues in cellular network is the design of user tracking strategy. Since users in a cellular network are mobiles and they could be anywhere in the network, the network has to determine the location of users to a minimum of cell granularity for service delivery. The process that the network searches for mobile user is called terminal paging. Upon a service arrival, the network broadcasts a paging message in the area where the mobile is expected to be found. The mobile listens to the paging message and sends a response to the network. Since the network has to determine the current location of the mobile user within fixed time delay constraint, the paging area is usually a cluster of neighboring cells around the last known location of the user. To prevent the paging area from becoming indefinitely large, each mobile reports its location to the network from time to time, and the process of reporting location is called location update or location registration. The heart of the design is the paging/registration optimization.

In the second and the third generation mobile cellular networks, the service area is partitioned into a number of location areas (LAs) [46]. Each LA contains a group of neighboring cells, and the location update is triggered when mobile moves from one LA to another. When there is call arrival for a mobile, the network broadcasts a paging message in all the cells in the current LA of the mobile; therefore, the network can find the location of a mobile in a single step. The problem with LA scheme is that LAs are not optimized on a user basis although the call arrival rate and the mobility of each individual user may be quite different. Current research efforts on the user tracking strategy for next generation wireless network is concentrated on user based location updating.

In the dynamic LA scheme, the size of LA for each user is dynamically determined based on the mobility and the call arrival rate of the user. The size of LA is chosen to minimize the paging/registration cost for each individual user under a fluid flow mobility model in [14]. In the dynamic location area assignment algorithm [47], the overlapping region of the new and the old LA is dynamically adjusted so that the expected dwell time of the mobile in the new LA is maximized.

In the timer-based location update scheme, mobile updates its location periodically. In the adaptive threshold scheme [48], the timer value is varied according to the current signaling load on the reverse control channel. The performance of the scheme is analyzed under one-dimensional random walk mobility assumption. The result shows that the adaptive threshold scheme performs better than the static timer-based scheme in which the timer value is fixed. An analytical model is used in [49] to study optimal timer value that minimizes the paging/registration cost of individual user. Under a Gaussian user location distribution and Poisson call arrival assumption, it is shown that the timer-based scheme performs significantly better than the LA-based scheme [46].

In the movement-based location update scheme, each mobile counts the number of cell boundary crossings. The location update is triggered when the counter exceeds a predetermined threshold. The threshold value can be selected dynamically for each individual user based on the user's call arrival rate and movement pattern. In the dynamic movement-based location update scheme [50], an analytical model is used to study optimal movement threshold that minimizes the paging/registration cost of individual user under a general cell residence time distribution and symmetric random walk movement assumption.

In the distance-based location update scheme, each mobile keeps track of the distance from the location where it performed the last location update. When this distance exceeds a threshold, a new location update is triggered. Dynamic programming is used in [38] to determine the distance threshold that minimizes signaling cost under one-dimensional symmetric random walk mobility model. An iterative method in [39] to determine the distance threshold under two-dimensional symmetric random walk mobility assumption. In the comparative study [40], the distance-based scheme is compared with the timer-based and the movement-based schemes under memoryless movement patterns and the movements with Markovian memory on a ring topology. It is shown that the signaling cost of the distance-based scheme is lower than the signaling cost of the timer-based and the movement-based schemes.

In the predictive distance-based mobility management scheme [36], the mobile updates its location as well as its velocity (i.e., speed and direction) during the location update process. Since the network can predict the location of mobile using the mobile's speed and direction, mobile updates its location when the distance between the current location and the predicted location exceeds a predetermined threshold. The network determines the probability distribution function of the mobile's location using the last report on location and velocity. Upon a call arrival, the network pages mobile at around the predicted location in the order of descending probability.

3.2 Mobility Model

The mobility models introduced in [38][39][40][50] consider individual user mobility; however, the movement patterns of users are assumed to be statistically symmetrical in all possible directions of movement; therefore, users are not biased in direction. In wireless networks, it is often assumed that the trajectory of user is random for the convenience of analysis, but mobile users have their own reason to move from one place to another, specially when they move with high speed. They usually take the shortest path to their destination, and their trajectory is highly predictable. It is obvious that the signaling cost for user tracking can be reduced by employing some predictive mechanism for those mobiles having highly predictable trajectory as in [36]. However, is it cost effective to predict the location of mobile users if their trajectory is highly erratic? The answer depends on how the network predicts the location of mobiles and the level of uncertainty in mobiles' trajectory. To answer the question, a mobility model that captures both the predictive and the non-predictive aspects of user mobility should be used in the analysis. In this chapter, 2-D Brownian motion with drift, $\mathrm{B}(\mu, \sigma)$ [43], is used to model user mobility.

3.3 Mobility Classes and Location Updating

A meandering pedestrian walking in the street does not need to update its location as long as the mobile remains in a small local area. In the following, such mobiles are referred as *class 1* mobiles. The analytical definition of *class 1* mobile will follow later when dynamic mobility classification is introduced. A mobile on a highly directed path such as train or vehicle on highway travels long distance without changing its speed and direction frequently. Such mobile does not need to update its location as long as it maintains its speed and direction, and they are referred as *class 2* mobiles. The analytical definition of *class 2* mobile will follow later. Assuming that each mobile is capable of determining its position, speed and direction, the location update schemes for *class 1* and *class 2* mobiles are presented as follows.

3.3.1 Non-predictive Distance-based Location Update Scheme for Class 1 Mobiles

Consider the trajectory of a mobile in Figure 2.1. The mobile changes its speed and direction frequently, if the mobile updates the changes in its speed and direction, the mobile will generate a lot of update messages while it stays in the same local area. In the non-predictive distance-based location update scheme, a *class 1* mobile reports its position when the distance between the current position and the last updated position becomes r where r is the distance threshold that triggers position update. The distance threshold should be optimized based on the user's call arrival pattern and mobility.

3.3.2 Predictive Distance-based Location Update Scheme for Class 2 Mobiles

Consider the trajectory of a mobile in Figure 2.2. Assuming that the mobile updated its mobility status (i.e., position, speed and direction) at time t_0, the network can predict the position of the mobile at time t ($> t_0$) as

$$p_t = p_0 + v_0 \cdot (t - t_0), \tag{3.1}$$

where p_0 is the position of the mobile at time t_0, and v_0 is the velocity (i.e., speed and direction) of the mobile at time t_0. In Figure 2.2, the mobile changes its speed and direction at p_1, and the distance between the mobile's position and the predicted position begins to diverge. Since the position prediction is based on the last mobility status update, it is assumed that the mobile is aware of the predicted position. The mobile updates its mobility status again when the distance between the current position and the predicted position becomes r where r is the distance threshold that triggers mobility status update. The distance threshold should be optimized based on the user's call arrival pattern and mobility.

3.3.3 Terminal Paging

When a call arrives for a *class 1* mobile, the network broadcasts paging message in the cells that overlap with any part of a circular region of radius r, which is centered at the last reported position of the mobile. For a *class 2* mobile, the network broadcasts paging message in the cells that overlap with a circular region of radius r, which is centered at the predicted position of the mobile. Therefore, the network can find the target mobile in a single step. The distance threshold dictates the size of the paging area, and it is subject to optimization with respect to the mobility and the call arrival rate of each individual user.

3.3.4 Dynamic Mobility Classification

In the later sections, the analysis shows that the predictive distance-based location update scheme performs better than the non-predictive distance-based location update scheme in terms of update frequency if the trajectory of mobile is highly predictable. If the trajectory of mobile is highly erratic, the non-predictive distance-based location update scheme performs better than the predictive distance-based location update scheme. Based on this result, a predictability-based adaptive location update scheme is proposed, in which each mobile is classified to either *class 1* or *class 2* according to the predictability of its trajectory. *Class 1* mobiles are prescribed the non-predictive distance-based location update scheme, and *class 2* mobiles are prescribed the predictive distance-based location update scheme. In stead of measuring the predictability of mobile's trajectory explicitly, the network operation cost (i.e., paging and location update cost) is used as a criterion for mobility classification. A mobile is classified to *class 1* if the signaling cost of the non-predictive distance-based location update scheme is smaller than the signaling cost of the predictive distance-based location update scheme. Otherwise, the mobile is classified to *class 2*.

Since mobility classification requires computation of the signaling cost for the predictive and the non-predictive location update schemes, each mobile needs to determine its own mobility parameters. If the user mobility is stationary (i.e., statistical properties of movement pattern doesn't change in time), each mobile can estimate its mobility parameters accurately by monitoring its own movement over a period of time. However, in reality the user mobility is not stationary, and the very idea behind the mobility classification is to exploit the non-stationarity of the user mobility. For example, when a mobile user is doing random walk in a department store, the mobile should choose the non-predictive distance-based location update scheme. If the user decides to take a taxi and starts moving on the highway, the mobile should detect the changes in the motion pattern and choose the predictive distance-based location update scheme. Therefore, the mobile should be able to detect the changes in mobility fast enough to take advantage of mobility classification. If the mobile doesn't detect the changes and uses the non-predictive scheme while the user is on the highway or the predictive scheme while the user is doing random walk, then the performance of the signaling cost for location tracking will be large. Therefore, the duration of movement history, on which the mobility parameters are measured, should be short enough to detect the changes in movement pattern rather quickly. On the other hand, mobile needs to determine the signaling cost to choose appropriate location update scheme, and the calculation of signaling cost requires estimations on the mobility parameters. Therefore, the duration of movement history, in which the mobility parameters are measured, should be long enough to result accurate mobility parameter estimation.

From a practical stand point, the mobility classification may not require such complex computation or mobility modeling. For example, if a mobile is moving with speed 60 MPH or more, one may know almost for sure that the mobile is on a highly directed path. The mobile may choose the predictive distance-based location update scheme. On the other hand, a slow mobile may choose the non-predictive distance-based location update scheme regardless of the "predictability" of its trajectory since the performance difference between the predictive and the non-predictive scheme is small when the speed of mobile is low.

Other than the complexity related to mobility classification, the non-predictive distance-based location update scheme for *class 1* mobile is essentially same as the conventional distance-based location update scheme [38][39][40], and the predictive distance-based location update scheme for *class 2* mobile is similar to the predictive distance-based mobility management scheme [36]. The only difference is that in [36] the network determines the probability distribution function (PDF) of mobile's location using the last report on location and velocity (i.e., speed and direction), and pages the mobile around the predicted location in the order of descending probability. In our scheme, the network pages mobile around the predicted location in a single step, and there is no need to calculate location PDFs.

3.4 Performance Analysis

In the following performance analysis, call arrivals are assumed to be independent of user movement. The call arrival process is assumed to be Poisson, and the call inter-arrival time is exponentially distributed as

$$p_{call}(t) = \lambda \cdot \mathrm{e}^{-\lambda t} \tag{3.2}$$

where λ is the average call arrival rate.

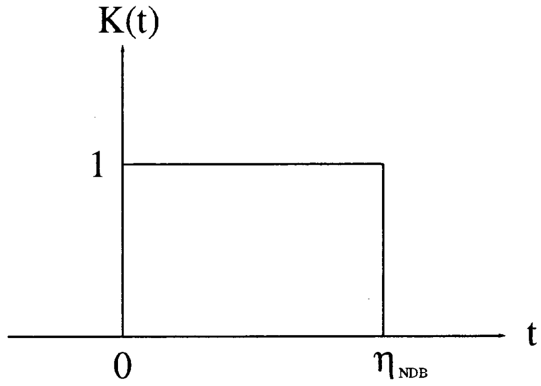


Figure 3.1 Indicator function.

3.4.1 Update Frequency of Non-Predictive Distance-based Location Update Scheme

There are three types of location update events, namely, update upon call arrival, update caused by mobile that initiates call and update triggered by distance measure. For simplicity, the location updates caused by initiating calls are ignored. In the non-predictive distance-based location update scheme (NDB), mobile updates its location either when there is call arrival or when the distance between the current position and the last updated position becomes a predetermined threshold. Using the memoryless property of the Poisson call arrival process, the time between location updates is expressed as $T_{NDB} = min(M,\ T_{call})$ where T_{call} is a random variable representing call inter-arrival time, and M is the dwell time of the mobile in the circular region of radius r centered at the initial position. Assume that the initial position of the mobile is $(0,0)$ at time $t = 0$. An indicator function is defined as follows.

$$K(t)=\begin{cases}1 & \text{if } T_{NDB} > t,\\ 0 & \text{otherwise}\end{cases} \tag{3.3}$$

The indicator function is shown in Figure 3.1. The area under the curve is equal to the random variable T_{NDB}; therefore,

$$T_{NDB} = \int_0^{\infty} K(t)\ dt. \tag{3.4}$$

Taking the average on the both side,

$$
\begin{aligned}
E\{T_{NDB}\} &= E\left\{\int_0^{\infty} K(t)\, dt\right\} \\
&= \int_0^{\infty} E\{K(t)\}\, dt \\
&= \int_0^{\infty} \Pr\{T_{NDB} > t\}\, dt \\
&= \int_0^{\infty} \Pr\{M > t,\ T_{call} > t\}\, dt
\end{aligned}
\tag{3.5}
$$

Since user movements and call arrivals are assumed independent,

$$
\begin{aligned}
E\{T_{NDB}\} &= \int_0^{\infty} \Pr\{M > t\} \cdot \Pr\{T_{call} > t\}\, dt \\
&= \int_0^{\infty} [1 - \Pr\{M < t\}] \cdot \mathrm{e}^{-\lambda t}\, dt \\
&= \frac{1}{\lambda} - \int_0^{\infty} \Pr\{M < t\} \cdot \mathrm{e}^{-\lambda t}\, dt
\end{aligned}
\tag{3.6}
$$

where $\Pr\{M < t\}$ is the cumulative probability distribution of hitting time M, and $\Pr\{T_{call} > t\}$ is the probability that call inter-arrival time is greater than t. By recognizing the second term in (3.6) as Laplace transform of the cumulative distribution and using the following definition of Laplace transform,

$$
\mathcal{L}\{\Pr\{M < t\}\} = \frac{\phi_M(s)}{s}, \tag{3.7}
$$

where $\phi_M(s)$ is Laplace transform of the dwell time distribution (2.2). (3.6) is expressed as

$$
\begin{aligned}
E\{T_{NDB}\} &= \frac{1}{\lambda} - \frac{\phi_M(\lambda)}{\lambda} \\
&= \frac{1}{\lambda}[1 - \phi_M(\lambda)]
\end{aligned}
\tag{3.8}
$$

Defining the inverse of the average time between location updates as location update frequency, the location update frequency of the non-predictive distance-based location update scheme becomes

$$
f_{NDB}(\lambda, r, |\mu|, \sigma) = \frac{\lambda}{1 - \phi_M(\lambda)}. \tag{3.9}
$$

The asymptotic behavior of the update frequency is determined as follows. If the call arrival rate λ approaches infinity, $\phi_M(\lambda)$ approaches zero; therefore,

$$\lim_{\lambda \to \infty} f_{NDB}(\lambda, r, |\mu|, \sigma) = \infty. \tag{3.10}$$

When the call arrival rate is large, the location update is mostly triggered by call arrivals, and the update frequency will be close to call arrival rate. If the call arrival rate λ approaches zero, then $\phi_M(\lambda)$ approaches one, and both the denominator and the numerator in (3.9) approaches zero. Using L'Hopital's rule,

$$\lim_{\lambda \to 0} f_{NDB}(\lambda, r, |\mu|, \sigma) = \frac{1}{[-\frac{d}{d\lambda}\, \Phi_M(\lambda)]_{\lambda=0}}. \tag{3.11}$$

From the moment theorem, the denominator in (3.11) is the first moment of the hitting time distribution. (3.11) can be expressed as

$$\lim_{\lambda \to 0} f_{NDB}(\lambda, r, |\mu|, \sigma) = \frac{1}{E\{M\}} \tag{3.12}$$

where

$$E\{M\} = \frac{r \cdot I_1(|\mu| r/\sigma^2)}{|\mu| \cdot I_0(|\mu| r/\sigma^2)}. \tag{3.13}$$

In other words, when the call arrival rate is zero, the location update is triggered by distance measure only, that is, the update frequency is the inverse of the average hitting time in (3.13). If the distance threshold r approaches infinity, $\phi_M(\lambda)$ approaches zero, and

$$\lim_{r \to \infty} f_{NDB}(\lambda, r, |\mu|, \sigma) = \lambda, \tag{3.14}$$

that is, the location update is triggered mostly by the call arrivals, and the update frequency is equal to call arrival rate λ. If the distance threshold r approaches zero, then $\phi_M(\lambda)$ approaches one, and

$$\lim_{r \to 0} f_{NDB}(\lambda, r, |\mu|, \sigma) = \infty. \tag{3.15}$$

Since mobile updates its location whenever the distance measure exceeds the distance threshold, the update frequency becomes infinity as the distance threshold approaches zero.

3.4.2 Update Frequency of Predictive Distance-based Location Update Scheme

In the predictive distance-based location update scheme, mobile updates its mobility status when the distance between the current position and the predicted position exceeds distance threshold. It is assumed that mobile determines its velocity (i.e., speed and direction) by monitoring its displacement during the last τ units of time before location update. Consider two mobiles j and k. They are located at the same position p_0 at time t_0. Assume that the movement of mobile j is described by Brownian motion, $\mathrm{B}(\mu_j \neq 0, \sigma_j \neq 0)$, while mobile k is moving with constant velocity, $\mu_k \neq 0$. Also assume that mobile j updated its position, speed and direction (i.e., p_0, $|V|$, $\angle V$) at time t_0. Mobile j determined its velocity, V, by monitoring its displacement during interval $[-\tau, 0]$. Since the displacement of mobile j during interval $[-\tau, 0]$ has Gaussian distribution with mean $\mu_j \tau$ and variance ${\sigma_j}^2 \tau$, V has Gaussian distribution with mean μ_j and variance ${\sigma_j}^2/\tau$. If the velocity of mobile k is equal to V (i.e., $\mu_k = V$), then the predicted position of mobile j at time t ($> t_0$) is the position of mobile k at time t. Then mobile j will update its mobility status again when the distance between its current position and the predicted position (i.e., the distance between mobile j and k) exceeds distance threshold r. The relative motion of mobile j with respect to mobile k is 2-D Brownian motion $\mathrm{B}(\mu_j - V, \sigma_j)$, and the time at which the distance between mobile j and mobile k becomes r is the hitting time of the Brownian motion $\mathrm{B}(\mu_j - V, \sigma_j)$ where the radius of the boundary is r. In the following, $\mu_j - V$ is referred as velocity jitter of mobile j at time t_0 and denoted it as J. Using (2.2) and (3.8), the average time between location updates (i.e., mobility status update) in the predictive distance-based location update scheme (PDB) becomes a function of velocity jitter J.

$$E\{T_{PDB}|J\} = \frac{1}{\lambda}\left[1 - \frac{I_0(|J|r/\sigma^2)}{I_0(\sqrt{2\lambda + |J|^2/\sigma^2}r/\sigma)}\right] \tag{3.16}$$

Since V is Gaussian distributed with mean μ_j and variance $\sigma_j{}^2/\tau$, the velocity jitter J is Gaussian distributed with zero mean and variance $\sigma_j{}^2/\tau$. Therefore, the amplitude of velocity jitter is Rayleigh distributed [45]:

$$p_Y(y, \sigma, \tau) = \frac{y}{\sigma^2/\tau} \cdot \mathrm{e}^{-\frac{y^2}{2\sigma^2/\tau}} \tag{3.17}$$

where Y=$|J|$. Taking the average of (3.16), the average time between location updates in the predictive distance-based location update scheme becomes

$$E\{T_{PDB}\} = \frac{1}{\lambda} - \int_0^\infty \frac{I_0(yr/\sigma^2) \cdot p_Y(y, \sigma, \tau)}{\lambda \cdot I_0(\sqrt{2\lambda + y^2/\sigma^2}r/\sigma)} dy \tag{3.18}$$

where

- r (km): distance threshold,
- σ^2 (km^2/hour): variance parameter of the Brownian motion,
- τ (hours): duration of velocity monitoring interval.

By defining the inverse of the average time between location updates as location update frequency, the location update frequency of the predictive distance-based location update scheme becomes

$$f_{PDB}(\lambda, r, \sigma, \tau) = \frac{\lambda}{1 - \int_0^\infty \frac{I_0(yr/\sigma^2) \cdot p_Y(y, \sigma, \tau)}{I_0(\sqrt{2\lambda + y^2/\sigma^2}r/\sigma)} dy}. \tag{3.19}$$

Notice that the location update frequency above is independent of drift velocity.

Except for $\lambda = 0$, the update frequency of the predictive distance-based location update scheme has similar asymptotic behavior as the non-predictive distance-based location update scheme. For $\lambda = 0$, applying L'Hopital's rule to (3.19) results

$$\lim_{\lambda \to 0} f_{PDB}(\lambda, r, \sigma, \tau) = \frac{1}{\int_0^{\infty} \frac{r \cdot I_1(yr/\sigma^2) \cdot p_Y(y, \sigma, \tau)}{y \cdot I_0(yr/\sigma^2)} dy}. \tag{3.20}$$

Using (3.13), (3.20) becomes

$$\lim_{\lambda \to 0} f_{PDB}(\lambda, r, \sigma, \tau) = \frac{1}{E\{E\{M|\mu = J\}\}}. \tag{3.21}$$

When the call arrival rate is zero, the location update is triggered by distance measure only.

3.4.3 Signaling Cost

The size of paging area is subject to optimization due to the trade off between the signaling cost of terminal paging and the signaling cost of location update. The size of paging area also affects the utilization of other resources such as signaling and switching in the wired backbone; however, in the following analysis the radio bandwidth is considered as a primary resource and regarded as a cost. Signaling cost (bits/user/hour) is formulated as follows.

$$Cost = n \cdot \lambda \cdot w_p + f \cdot w_u \tag{3.22}$$

where

- n is the number of cells in paging area,
- λ (calls/user/hour) is the average call arrival rate,
- w_p (bits/call/cell) is the radio bandwidth needed for terminal paging,
- w_u (bits/update) is the radio bandwidth needed for location update,

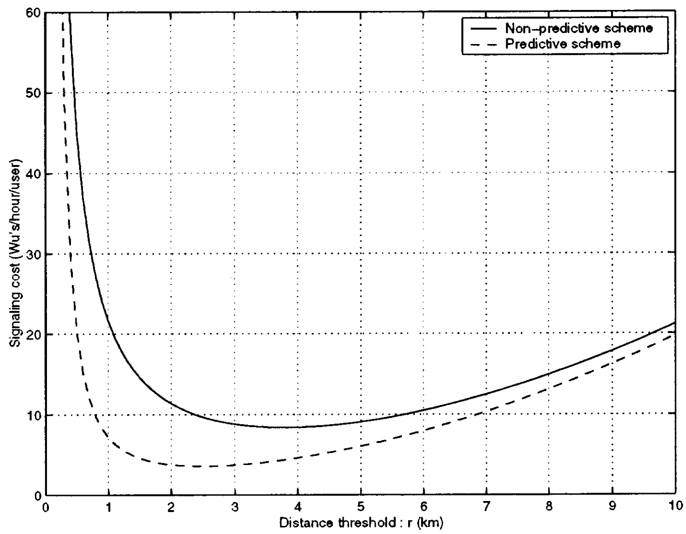


Figure 3.2 Normalized signaling cost vs. radius of paging area.

- f (updates/user/hour) is the location update frequency.

Assuming that cells are small, the number of cells which overlap with any part of a circular region of radius r can be approximated as

$$n \simeq \frac{\pi r^2}{b} \tag{3.23}$$

where b (km^2) is the size of cell. For simplicity, the signaling cost is normalized by w_u. The normalized signaling cost becomes

$$c = \pi r^2 \lambda \epsilon + f \tag{3.24}$$

where $c = Cost/w_u$ and $\epsilon = (w_p/b)/w_u$. For λ=0.6, $|\mu|$=20, σ^2=2, τ=0.1 and ϵ=0.1, the normalized signaling cost versus distance threshold is plotted in Figure 3.2. Figure 3.2 shows the trade off between the cost of terminal paging and the cost of location update. When r is increased, the cost of terminal paging increases while the cost of location update decreases. On the other hand, if r is decreased, the cost of the location update increases while the cost of terminal paging decreases.

Since the paging cost is increasing function of r and the location update cost is decreasing function of r, there exists finite r that minimizes the overall signaling

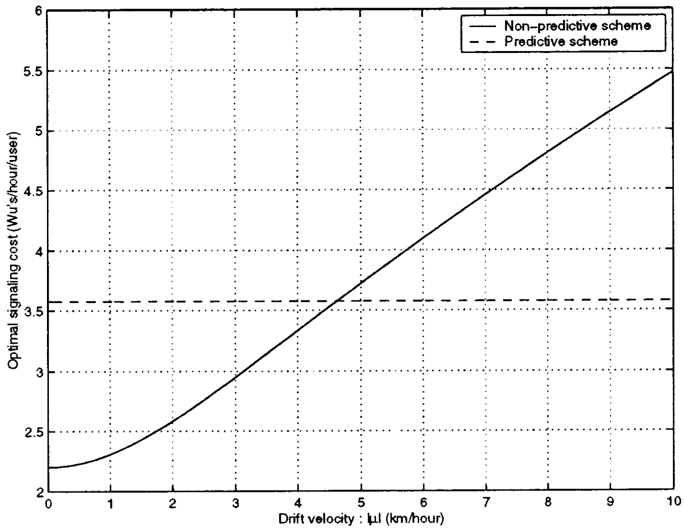


Figure 3.3 Optimal signaling cost vs. drift velocity.

cost. One can numerically evaluate the optimal distance threshold for each scheme, namely r^*_{NDB} for the non-predictive scheme and r^*_{PDB} for the predictive scheme. The optimal signaling cost for each scheme becomes

$$c^*_{NDB}(\lambda, |\mu|, \sigma) = \pi r^{*2}_{NDB}\lambda\epsilon + f_{NDB}(\lambda, r^*_{NDB}, |\mu|, \sigma), \tag{3.25}$$

$$c^*_{PDB}(\lambda, \sigma, \tau) = \pi r^{*2}_{PDB}\lambda\epsilon + f_{PDB}(\lambda, r^*_{PDB}, \sigma, \tau). \tag{3.26}$$

For λ=0.6, σ^2=2, τ=0.1 and ϵ=0.1, the optimal signaling cost versus drift velocity is plotted in Figure 3.3. The solid curve indicates the signaling cost of the non-predictive scheme, and the dotted line indicates the signaling cost of the predictive scheme. The signaling cost of the non-predictive scheme increases as the drift velocity increases. The signaling cost of the predictive scheme is constant regardless of the drift velocity. When $|\mu|$ <4.62, the signaling cost of the non-predictive scheme is smaller than the signaling cost of the predictive scheme. When $|\mu|$ >4.62, the predictive scheme is more cost efficient than the non-predictive scheme.

3.4.4 Performance of Dynamic Mobility Classification

In the previous subsection, it is shown that the non-predictive distance-based location update scheme is more cost efficient than the predictive distance-based location update scheme when the trajectory of mobile is highly random (i.e., $|\mu| \ll \sigma^2$). When the mobile is highly biased in direction (i.e., $|\mu| \gg \sigma^2$), the predictive distance-based location update scheme results lower signaling cost. Therefore, the combination of the predictive and the non-predictive schemes can reduce the overall signaling cost by choosing an appropriate update scheme dynamically. In the dynamic classification, a mobile is classified to *class 1* if $c^*_{NDB}(\lambda, |\mu|, \sigma) < c^*_{PDB}(\lambda, \sigma, \tau)$, otherwise, the mobile is classified to *class 2*. The performance of the hybrid location update scheme that uses the dynamic mobility classification is compared with the performance of the predictive and the non-predictive distance-based location update schemes. The service area of the cellular network is assumed to be two dimensional and infinite in size. Two types of users are considered, namely, pedestrian users and vehicular users. The movements of pedestrian and vehicular users are modeled by Brownian motion $\mathrm{B}(\mu_p, \sigma_p)$ and $\mathrm{B}(\mu_v, \sigma_v)$ respectively. It is assumed that the velocity monitoring interval is τ=0.1 (i.e., 6 minutes), and the ratio between w_p/b and w_u is ϵ=0.1 as before.

In Figure 3.4, the signaling cost versus call arrival rate is plotted for $|\mu_p|$=0, ${\sigma_p}^2$=0.1, $|\mu_v|$=10 and ${\sigma_v}^2$=2. It is assumed that each type of user contributes 50% of the total user population in the network. Figure 3.4 shows that the signaling cost of the proposed scheme is smaller than the signaling cost of the predictive and the non-predictive distance-based location update schemes for call arrival rate between 0.1 and 1 (calls/hour/user). The difference in signaling cost increases as the average call arrival rate increases. When the drift velocity of vehicular users is increased, the signaling cost of the non-predictive scheme increases while the signaling costs of

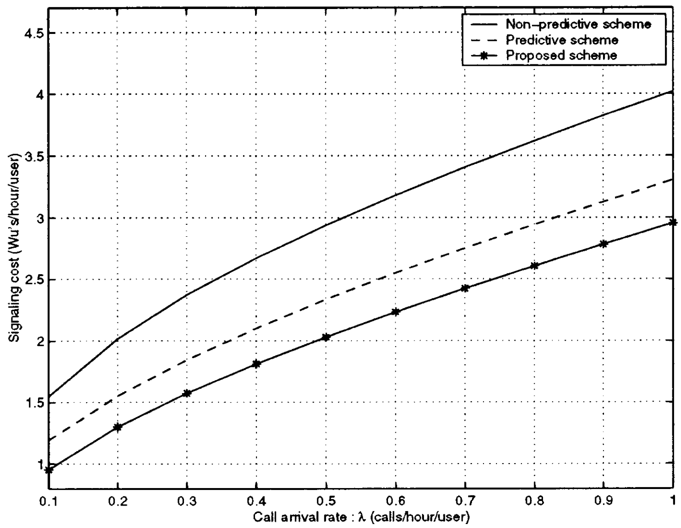


Figure 3.4 Signaling cost vs. call arrival rate.

the proposed scheme and the predictive scheme remain constant. It is because the location update frequency of *class 2* mobile (i.e., vehicular user) is independent of drift velocity in both the proposed scheme and the predictive scheme. When the variance parameter of pedestrian users is decreased, the signaling cost for all three schemes decreases; however, the difference between the signaling cost of the proposed scheme and the signaling cost of the predictive scheme increases. In other words, the signaling cost of the proposed scheme decreases more rapidly than the signaling cost of the predictive scheme. For λ=0.6, the signaling cost versus percent of the vehicular user population is plotted in Figure 3.5. The mobility of each user type is same as before, and the population of the vehicular users are varied from 10% to 90%. As the population of vehicular user increases, the signaling cost of each scheme increases. When the vehicular users are less than 25% of total user population, the signaling cost of the non-predictive scheme is smaller than the signaling cost the predictive scheme. When vehicular users are more than 25%, the predictive scheme results lower signaling cost than the non-predictive scheme. In any case, the proposed scheme outperforms the predictive and the non-predictive schemes.

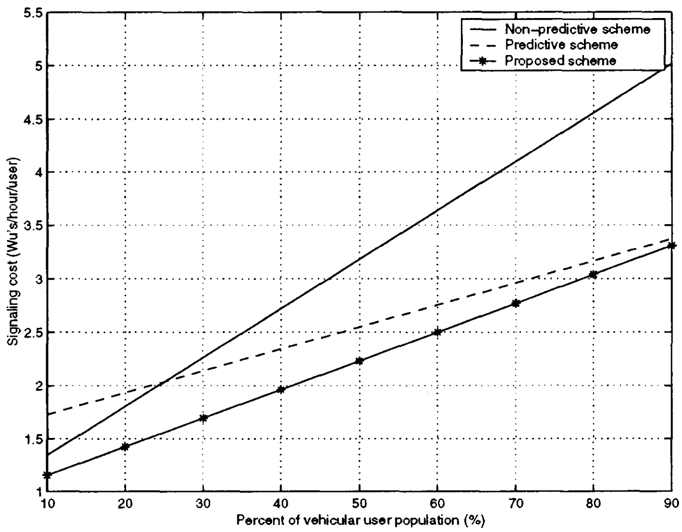


Figure 3.5 Optimal signaling cost vs. percent of vehicular users.

CHAPTER 4

LOCATION-BASED SERVICE (LBS) PROVISIONING FOR NEXT GENERATION WIRELESS NETWORKS

4.1 Introduction

As the service demand in wireless networks becomes more and more data oriented, wireless opens up a new dimension of services in which information is custom-tailored to user's location. An example of such location-aware application service is geographically-targeted message delivery service called geocasting. In geocasting, a user can send a message or establish communication with other mobile users in a given area of interest. The area of interest is referred as target area [51]. To deliver a message to the mobiles in the target area, the network needs to detect the presence of the mobiles in the target area. If the target area coincides with cellular layout then cellular base stations may broadcast paging messages in the corresponding cells to determine the presence of the target mobiles. If the shape of target area is arbitrary, the cell-base location management schemes, which is used in conventional mobile cellular networks, cannot provide sufficient information to determine whether a mobile is inside or outside of the target area.

Consider a target area in Figure 4.1. The target area is defined by a message source or an application service that wants to send a message to that area. In Figure 4.1, the target area is located in the middle of three cells. The cell that has any overlapping region with the target area is called embedding cell. Since a location sensitive message is intended for only those mobiles inside the target area, one way to deliver a message from the application service is to send the message itself and the target area boundary information to all the mobiles in the embedding cells. Since it is assumed that mobiles are equipped with self-geolocation capability, a mobile that receives the target area boundary information can determine its current geolocation.

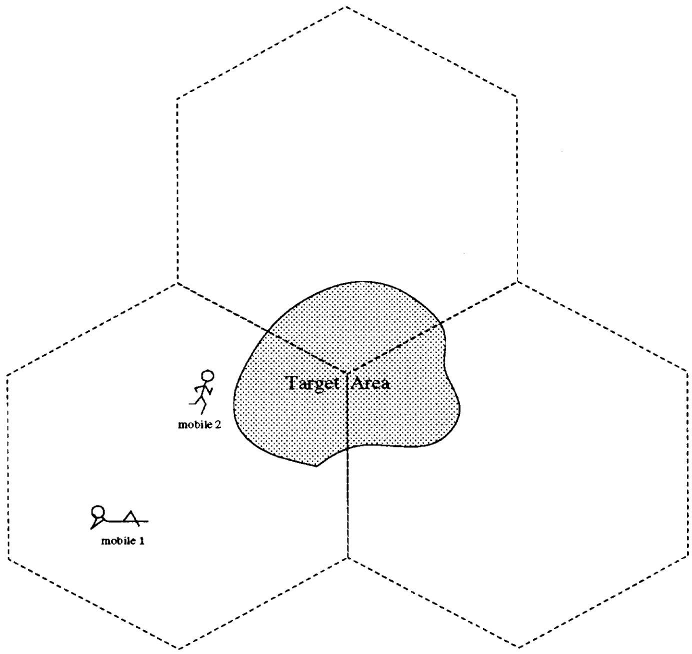


Figure 4.1 Target area of arbitrary shape.

If the mobile is inside of the target area, the mobile terminal may present the message from the application service to the user.

In a slightly different application which the application service wants to send a message to all the mobiles that ever visit the target area, the wireless network has to constantly monitor the embedding cells to detect the presence of new mobiles. Whenever the wireless network detects a new mobile in an embedding cell, the network may send the message from the application service and the target area information to the mobile. If the message source is sending a real-time stream data traffic to the target area, sending the stream traffic and the target area boundary information to all the mobiles in the embedding cells will consume significant amount of radio resource and battery power since the messages delivered to the mobiles outside the target area will be discarded. An alternative way to deliver a geolocation sensitive real-time message to a target area is to determine the geolocation of mobiles in the embedding cells before the network transmits the data traffic to mobile terminals.

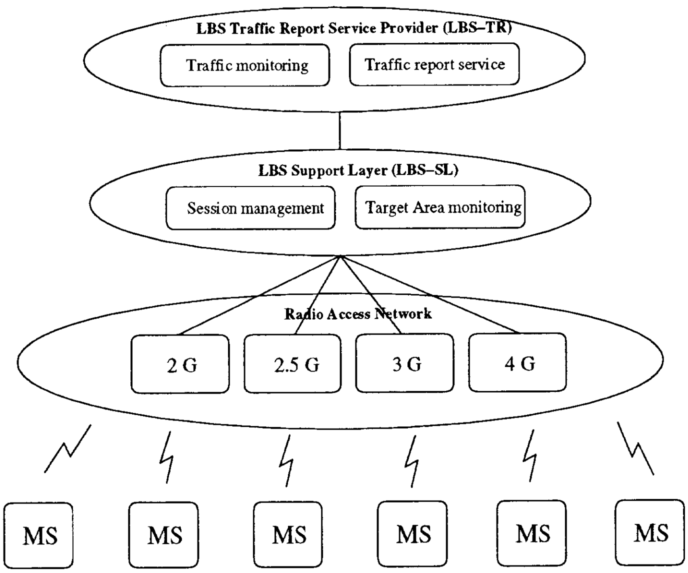


Figure 4.2 LBS-TR service architecture.

Each mobile in the embedding cells may report its geolocation to the network, and the network may inform the application service about the mobiles inside of the target area. If each mobile reports its geolocation frequently, then the network may detect all the mobiles that ever enter the target area. However, frequent geolocation update will consume significant amount of radio resource and battery power. Consider the two mobile users in Figure 4.1. Mobile 1 is not moving while mobile 2 is running toward the target area. If the both mobiles update their geolocation frequently, the network may be able to detect mobile 2 when it enters the target area. However, mobile 1 will waste its battery and radio resource. If they update their geolocation infrequently, then the wireless network may not detect mobile 2 while it pass through the target area.

Therefore, the geolocation update scheme of each individual mobile should be optimized based on its mobility and the QoS of the application service. In the following, a location-based traffic report service (LBS-TR) is presented as a case study. A geolocation update scheme is proposed for LBS-TR. In LBS-TR, the

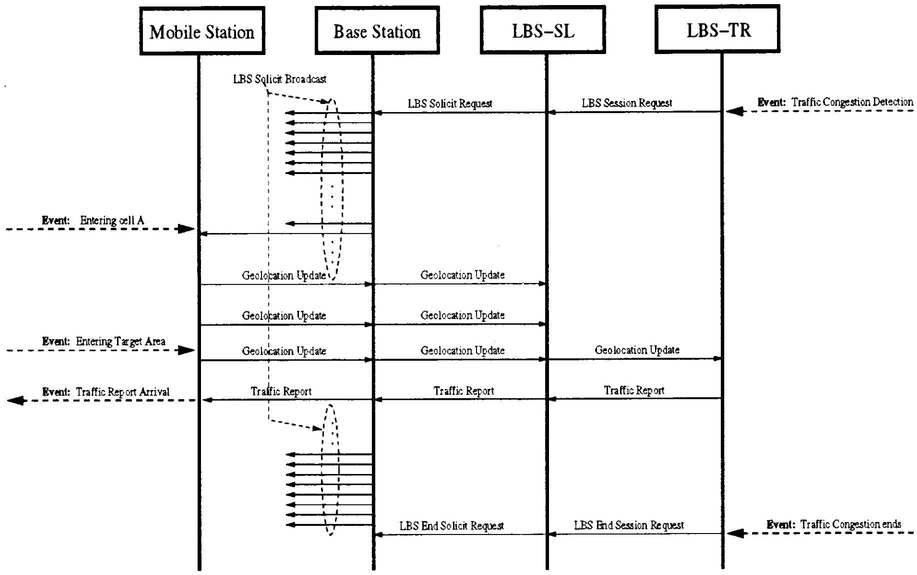


Figure 4.3 Signaling flow diagram for LBS-TR.

wireless network searches for mobile users who may find the traffic report message useful.

In other type of location-based services (LBS), the application service concerns about what is happening around a mobile user. The granularity of geolocation information of the mobile should be fine enough to differentiate the geographic relevance of the application service for the mobile but coarse enough not to cause too much signaling overhead in the network. For example, a navigation service for high speed mobile on the highway may require low resolution location tracking than a navigation service for mobiles wondering in a mall looking for a store. The wireless network may assume that the required resolution of the geolocation information is specified by the application service as a QoS requirement. The required resolution may change dynamically during a service. A location-based navigation service (LBS-NS) is presented as a case study. A hybrid geolocation update scheme is proposed for LBS-NS.

4.2 Case Study 1: Location-based Traffic Report Service (LBS-TR)

In the location-based traffic report service (LBS-TR), the service provider informs mobile users who are about to enter heavily jammed highway so that the users can avoid traffic congestion. LBS-TR service architecture is shown in Figure 4.2. The service provider monitors traffic conditions on all the major highways, and it maintains a data base for the potential target areas to which traffic report should be delivered in case of traffic congestion. When highway traffic congestion is detected, the service provider determines which target area it should send a traffic report and requests an LBS Session from the service support layer (LBS-SL). LBS-SL supports LBS application service by providing appropriate mobility information that is collected from the radio access network. LBS-SL does not need to know what kind of service is being provided by the service provider. The service provider sends the information about the target area (e.g., digital maps of the target area) to LBS-SL, and LBS-SL determines which radio access network is covering the target area. LBS-SL requests appropriate radio access networks to monitor the target area.

A simple signaling flow diagram illustrating LBS-TR service is shown in Figure 4.3. When the service provider detects highway traffic congestion, it maps the point of congestion to a set of target areas and sends a *LBS Session Request* message to LBS-SL. The *LBS Session Request* message includes information about the target areas. LBS-SL maps the target areas onto cellular layout and determines which access network has the coverage over the target areas. LBS-SL sends *LBS Solicit Request*s to appropriate radio access network, and the radio access network forwards the request to appropriate base stations. When the base stations receive *LBS Solicit Request*, they start broadcasting *LBS Solicit* message periodically. *LBS Solicit* message should include a signature of the application service (i.e., LBS-TR) so that only the subscribing mobiles can respond to the *LBS Solicit* message. Subscribing mobile

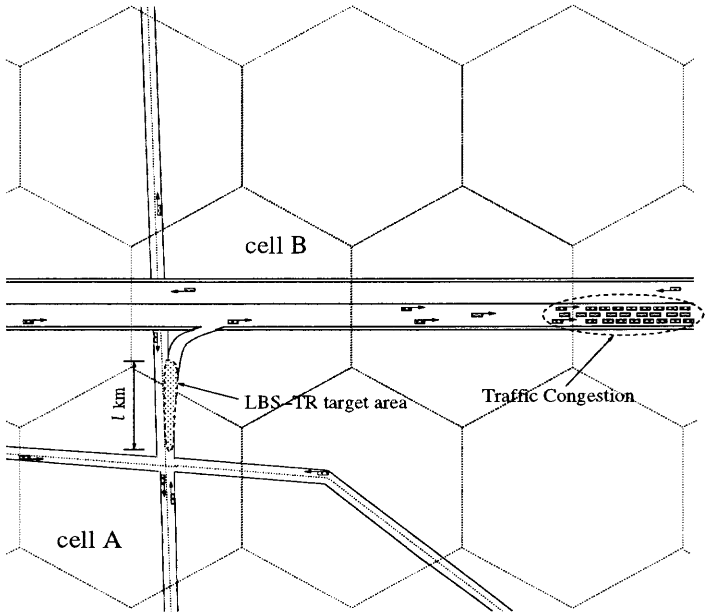


Figure 4.4 A target area of LBS-TR.

that receives *LBS Solicit* message starts self-geolocation and geolocation update. If the mobile moves to a non-embedding cell that doesn't have any overlapping region with the target area, the mobile will no longer hear the *LBS Solicit* message and stop both self-geolocation and geolocation update. The geolocation update messages received at the base stations are forwarded to LBS-SL. LBS-SL determines whether the mobile is inside the target area or not. If the mobile is in the target area, LBS-SL forwards the geolocation update message to the application service. If the mobile is not in the target area, LBS-SL discards the geolocation update message. Upon receiving the geolocation update message, the application service sends the mobile a *Traffic Report.* When the traffic congestion is over, the application service sends *LBS End Session Request* to LBS-SL, and LBS-SL forwards *LBS End Solicit Request* to the radio access network.

By mapping the target areas to cellular layout, the subscribing mobiles update their geolocation only when they are inside the embedding cells. The cell based self-geolocation triggering significantly improves the signaling scalability of LBS-TR.

However, the geolocation update frequency of each mobile inside the embedding cell still needs to be optimized. To determine the optimal geolocation update frequency of a mobile in the embedding cells, consider the target area shown in Figure 4.4.

Since mobile users who are not interested in traveling highways shouldn't be disturbed, the target area is designated in the north bound of the local street between the highway entrance and the local cross section. As shown in Figure 4.4, the target area is much smaller than the size of a radio cell, and the conventional location management schemes that keep track of mobile's location on cell level will not provide enough resolution. Assume that the distance between the highway entrance and the local cross section is ℓ km and that the local speed limit is 100 km/hour. A north bound vehicular user traveling with the speed limit will pass through the target area in $36 \times \ell$ seconds. Therefore, if geolocation update frequency is greater that $1/36\ell$ (updates/second), the wireless network will detect the mobile that enters the target area unless the mobile is speeding. Frequent geolocation update will guarantee the QoS for LBS-TR; however, if all the mobiles in cell A and B update their geolocation frequently, the overall signaling cost and battery consumption will be significant. If there is residential area or office complex where thousands of users sit tight, then most of the geolocation updates from these users are simply wasted. Therefore, the geolocation update frequency for each individual mobile needs to be optimized based its mobility, that is, slow mobiles may update their geolocation less frequently while high speed mobiles update their geolocation frequently.

Assume that the mobiles in the embedding cells in Figure 4.4 are moving with different speeds. Since the service is targeted for vehicular users traveling in the north bound of the local street, it is fair to assume that the target area in Figure 4.4 is one dimensional where the size of the target area is ℓ km. Since the size of the target area depends on the geography, and it doesn't usually change in time,

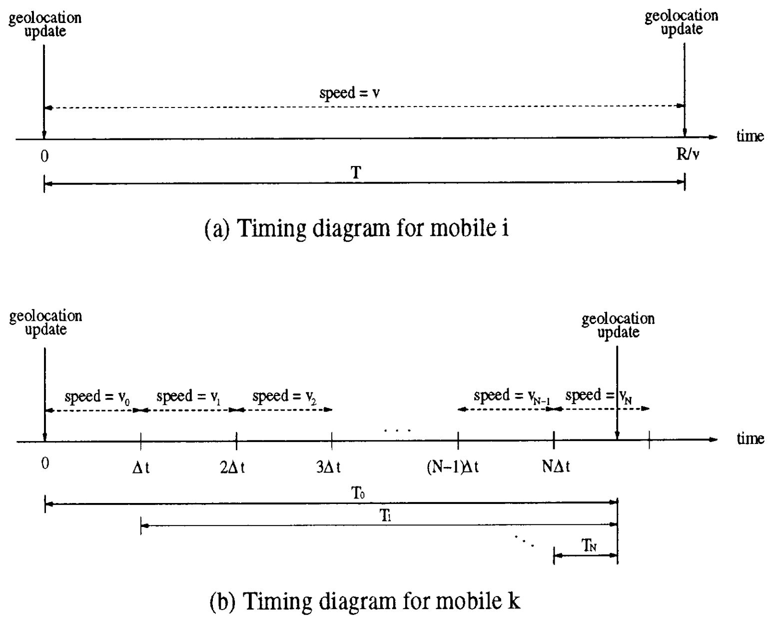


Figure 4.5 Timing diagram.

the information needed to calculate the optimal geolocation update frequency is the speed of mobile. The QoS of LBS-TR is satisfied if a mobile entering the target area updates its geolocation at least once before it moves out of the target area. If each mobile maintains its speed constant, the optimal geolocation update frequency is calculated by dividing the speed of mobile by the size of the one-dimensional target area. If the speed of the mobile is v km/hour, then the optimal geolocation update frequency for the mobile is

$$\begin{aligned} F &= \frac{\text{mobile speed}}{\text{size of target area}} \\ &= \frac{v}{\ell} \text{ (updates/hour)}, \end{aligned} \tag{4.1}$$

and the time between geolocation updates becomes

$$T = \frac{\ell}{v} \text{ (hours)}. \tag{4.2}$$

In other words, if the mobile updates its geolocation every ℓ/v (hours), the mobile will perform at least one geolocation update in the target area. If the speed of mobile is not constant, then the geolocation update timer value should be dynamically adjusted as the speed of mobile changes. In the following one dimensional analysis, it is shown that if the timer value is adjusted as the speed of mobile changes, the distance traveled by a mobile between two consecutive geolocation update events remains constant. In other words, the same optimality is achieved if a mobile updates its geolocation when the distance between its current position and the last updated position becomes ℓ.

Consider the timing diagram for mobile j and k in Figure 4.5. For ease of analysis, it is assumed that the movement of mobile is one dimensional. Both mobile j and k updated their geolocation at time $t = 0$. The speed of mobile j is v which is constant, and the time until the next geolocation update is $T = \ell/v$ (hours). Therefore, the distance traveled by mobile j between two consecutive geolocation updates is ℓ km. For mobile k, it is assumed that the speed changes periodically every Δt (hours). Let v_i be the speed of mobile k during time interval $(i \cdot \Delta t, (i+1) \cdot \Delta t)$ for $i = 0, 1, 2, \ldots$. The distance that mobile k travels during time interval $(i \cdot \Delta t, (i+1) \cdot \Delta t)$ is $d_i = v_i \cdot \Delta t$. The time until the next geolocation update, T_i, is calculated at each time instant $t = i \cdot \Delta t$. Since the speed of mobile k at time $t = 0$ is v_0, the time until next geolocation update, which is calculated at $t = 0$, is $T_0 = \ell/v_0$. At time $t = \Delta t$, T_1 is calculated using (4.2), but the size of target area is adjusted since the mobile has already moved d_0 (km) during the previous time interval $(0, \Delta t)$, that is,

$$T_1 = \frac{\ell - d_0}{v_1}. \tag{4.3}$$

Similarly, T_2, T_3, ... , T_i, ... are calculated as follows:

$$T_2 = \frac{\ell - d_0 - d_1}{v_2}, \ \ldots\ , T_i = \frac{\ell - \sum_{m=0}^{i-1} d_m}{v_i}, \ldots\ . \tag{4.4}$$

If there exists an integer N such that $0 \le T_N < \Delta t$, then the next geolocation update occurs during time interval $(N \cdot \Delta t, (N+1) \cdot \Delta t)$. The distance that mobile k moves during interval $(0, N \cdot \Delta t)$ is

$$d_{(0,N\cdot\Delta t)} = \sum_{m=0}^{N-1} d_m, \tag{4.5}$$

and the distance that the mobile moves during interval $(N \cdot \Delta t, N \cdot \Delta t + T_N)$ is

$$\begin{aligned} d_{(N\cdot\Delta t, N\cdot\Delta t+T_N)} &= T_N \cdot v_N \\ &= \left(\frac{\ell - \sum_{m=0}^{N-1} d_m}{v_N}\right) \cdot v_N \\ &= \ell - \sum_{m=0}^{N-1} d_m. \end{aligned} \tag{4.6}$$

Therefore, the distance that mobile k moves between two consecutive geolocation updates becomes

$$\begin{aligned} d_{(0,N\Delta t+T_N)} &= d_{(0,N\Delta t)} + d_{(N\Delta t,N\Delta t+T_N)} \\ &= \ell \ \ (km). \end{aligned} \tag{4.7}$$

The result can be generalized by letting Δt approach zero. For a mobile which changes its speed continuously, the distance traveled between two consecutive geolocation update events is always ℓ if the update timer is optimally adjusted as the speed of mobile changes.

From this simple one dimensional analysis, it is shown that the same QoS is achieved if geolocation update is triggered by the distance measure, that is, a mobile updates its geolocation when the distance between the mobile's current position and the last updated position becomes ℓ. This geolocation update strategy is known as distance-based geolocation updating. With the distance-based updating scheme, each mobile entering the target area updates its geolocation only once inside the target area.

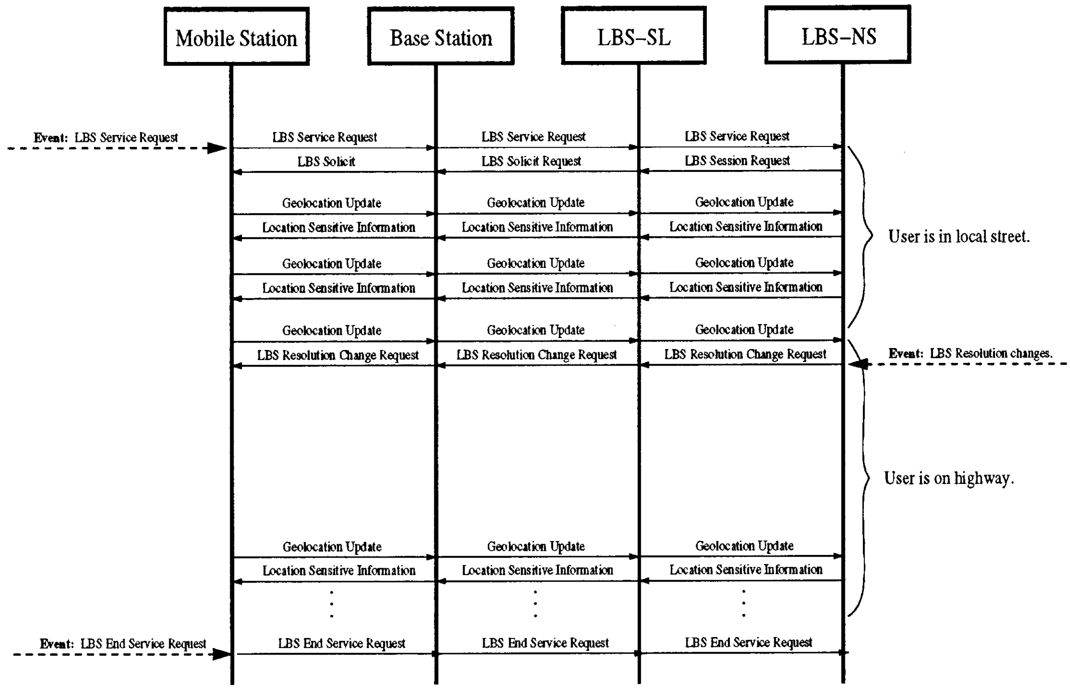


Figure 4.6 An exemplary signaling flow diagram for LBS-NS with dynamic resolution control.

4.3 Case Study 2: Location-based Navigation Service (LBS-NS)

Consider a location-based navigation service (LBS-NS) in which the user is given navigation instructions that are dynamically custom-tailored to the geolocation of the user. The resolution of navigation instructions should change as the user moves from highway to local street and from local street to building complex. The application service may require different granularity for the mobile's geolocation information as the user moves.

An exemplary signaling flow diagram for LBS-NS is shown in Figure 4.6. A mobile user may request a navigation service by sending a service request message to LBS-NS. LBS-NS initiates an LBS session by sending *LBS Session Request* to LBS Support Layer (LBS-SL). The *LBS Session Request* should include the QoS requirement such as the accuracy of mobile's geolocation information. LBS-SL requests the wireless network to keep track of the mobile's geolocation with a granularity specified by the application service. The granularity of geolocation information is the maximum difference between the mobile's true geolocation and the last reported geolocation. The granularity should be fine enough to differentiate the geographic relevance of surrounding landmarks so that LBS-NS can provide appropriate navigation instructions. Consider a mobile user who wants to buy shoes at a shopping mall. When the user is driving on the highway, the highway exits may serve as landmarks in the navigation instruction. The mobile doesn't need to update its geolocation frequently since the highway exits are usually far apart. As the user approaches the shopping mall, the user needs more detailed navigation instruction such as how to drive toward the mall's parking area or how to walk through the mall to fine the shoe store. Therefore, the required granularity of mobile's geolocation information becomes small as the mobile approaches the destination.

Even though the actual quality of navigation service also depends on the way that LBS-NS chooses landmarks for navigation instructions, in the following the QoS of LBS-NS is defined as the granularity of geolocation information. A hybrid geolocation update scheme that satisfy the QoS of LBS-NS is proposed. In the hybrid scheme, the non-predictive and the predictive distance-based geolocation update schemes are dynamically selected based on the user's mobility pattern. For the description of the the non-predictive and the predictive distance-based geolocation update schemes, see Chapter 1.

4.3.1 Hybrid Distance-based Geolocation Update Scheme

Analysis in Chapter 1 shows that the update frequency of the predictive distance-based geolocation update scheme is smaller than the update frequency of the non-predictive distance-based geolocation update scheme when the trajectory of mobile is highly predictable. When the trajectory of mobile is highly random, the update

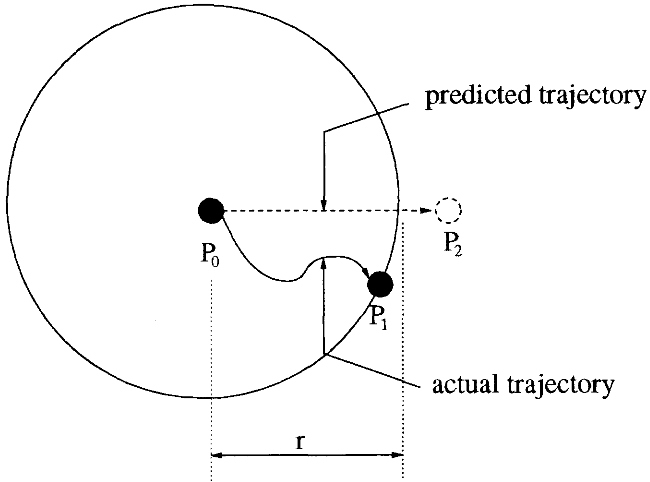


Figure 4.7 A mobile changing its geolocation update scheme from the non-predictive scheme to the predictive scheme.

frequency of the non-predictive scheme is smaller than the update frequency of the predictive scheme. Based on this analysis, a hybrid distance-based geolocation update scheme is proposed. In the proposed scheme, a mobile dynamically chooses the predictive or the non-predictive scheme whichever is expected to increase the time between two consecutive geolocation updates.

Consider the mobile in Figure 4.7. Assume that the mobile updated its geolocation at position P_0 at time t_0 using the non-predictive scheme. Also assume that the mobile arrives at position P_1 at time t_1. Even though the mobile updated its geolocation using the non-predictive scheme, the mobile can also determine its speed and direction at the time of geolocation update; therefore, the mobile "knows" what would've been the predicted position if the mobile had updated its geolocation using the predictive scheme. In Figure 4.7, the predicted position of the mobile at time t_1 is P_2. If the distance between P_1 and P_2 is less than the granularity of the geolocation information, r, then the mobile may "realize" that the predictive scheme would've performed better than the non-predictive scheme. The mobile changes the geolocation update scheme to the predictive scheme. Consider the mobile in Figure 4.8.

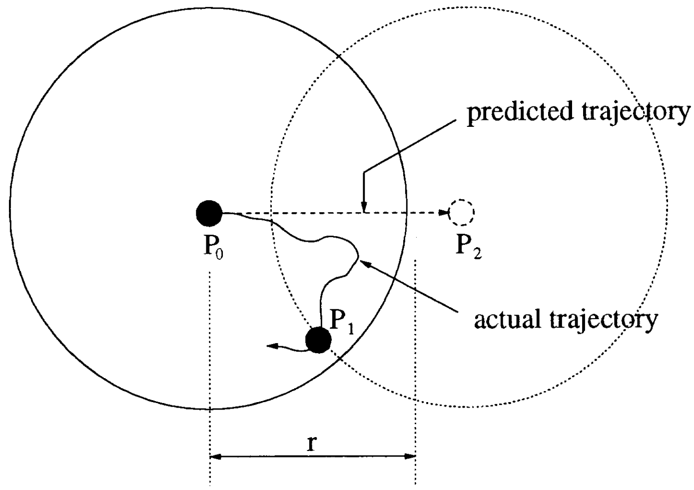


Figure 4.8 A mobile changing its geolocation update scheme from the predictive scheme to the non-predictive scheme.

Assume that the mobile updated its geolocation at position P_0 at time t_0 using the predictive scheme. By the time the mobile arrives at P_1, the distance between the mobile's position and the predicted position P_2 becomes r, and the mobile has to update its geolocation. Since the distance between the mobile's position (i.e., p_1) and the initial position (i.e., p_0) is smaller than r, the mobile may "realize" that the non-predictive scheme would've performed better than the predictive scheme and change the geolocation update scheme to the non-predictive scheme.

CHAPTER 5

MOBILITY-BASED ENERGY EFFICIENT MULTIHOP ROUTING FOR MULTIHOP CELLULAR NETWORKS

5.1 Introduction

Cellular ad hoc augmented network (CAHAN) architecture, which is a mixture of the mobile ad hoc network (MANET) and the cellular network, is proposed in [31]. Many variations of the proposed network architecture are possible depending on the set of functionalities inherited from the fully distributed self-organizing mobile ad hoc networks. One of the distinctive characteristics of CAHAN architecture is the peer-to-peer communication capability, that is, mobile stations can transmit messages to other mobile stations without the intervention of cellular base stations given that these mobiles are in close proximity. Using the peer-to-peer communication capability, the multihop cellular architecture can provide energy efficient multihop routes from mobiles to base stations. A simple illustration in Figure 5.1 shows how the energy consumption is reduced by multihop routing. In the figure, there are two mobile stations S and R and base station D. They are located on a straight line. From the simple radio path-loss model [52], the power of the received radio signal that is transmitted from S to D is inversely proportional to $(d_{S,D})^\alpha$ where $d_{S,D}$ is the distance between S and D, and α is the path-loss exponent. If the minimum received signal power required at each receiver for reliable reception of the transmitted data is P, then the required transmission power for a direct radio transmission from S to D is $P \cdot (d_{S,D})^\alpha$ assuming that the transmitter and the receiver antenna gains are unity. In the multihop route that S sends packets to R, and R relays the packets to D, the required transmission power at S is $P \cdot (d_{S,R})^\alpha$, and the required transmission power at R is $P \cdot (d_{R,D})^\alpha$ where $d_{S,R}$ is the distance between S and R, and $d_{R,D}$ is the distance between R and D. The multihop routing reduces total transmission

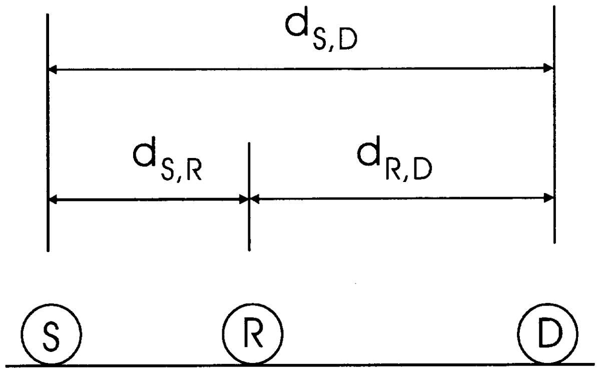


Figure 5.1 Multihop relay.

power[1], that is,

$$P \cdot (d_{S,D})^{\alpha} > P \cdot (d_{S,R})^{\alpha} + P \cdot (d_{R,D})^{\alpha}. \tag{5.1}$$

If there is more relay mobiles between S and D, the energy consumption will decrease even further.

While the multihop routing reduces the energy consumption, the capacity of a fixed multihop cellular network decreases as the density of mobile users increases. The capacity of an ad hoc network is analyzed in [53]. In the ad hoc network, the throughput for each user is limited not only by the capacity of a single radio link, but also the relay load imposed by other users. Since short range single-hop radio transmissions do not interfere with each other given that they are sufficiently distant, the number of possible simultaneous single-hop transmission increases linearly with the total area of the network if the density of the network is kept constant. Therefore, the capacity of the network, which is the total amount of data transferred per unit time in all possible simultaneous single-hop radio transmissions increases linearly with n where n is the number of users in the network. Assuming that the number of hops between each source and destination increases linearly with the square root of n, which is roughly the diameter of the network, the sum of the throughput for

[1]For simplicity, the power required at the receiver to process the received radio signal is ignored.

each communication session in the network is in the order of $n/\sqrt{n}$. Assuming that there are n such communication sessions in the network, the throughput available for each session is in the order of $1/\sqrt{n}$. A dramatic improvement on the capacity is suggested in [54] by taking advantage of user mobility in an ad hoc network. It is shown that the average long-term throughput per each communication session can be kept constant even when the number of users in the network increases.

In this chapter, a mobility-based energy efficient multihop routing protocol is proposed for up-link traffic, i.e., the data traffic from mobile stations to base stations. For down link traffic, the conventional single-hop radio transmissions is assumed to reduce energy consumption at mobile stations. In the proposed mobility-based multihop routing, the network takes advantage of user mobility to reduce the energy consumption. The proposed mobility-based multihop routing is illustrated in Figure 5.2. Assume that mobile S and R are moving along their trajectories that are indicated as dotted lines in the figure. If mobile S can predict how close it can get to base station D, then S may wait until it reaches the position where the distance to the base station is minimum, that is, where the required transmission power to send packets to the base station is minimum. In the following, this minimum distance to the base station is referred as minimum power distance (MPD) of the mobile. If mobile S has a neighboring mobile whose MPD is even smaller, S may relay its packets to the neighboring mobile. To reduce the total transmission energy, S relays packets to neighboring mobile R if

$$\delta_S^\alpha > (d_{S,R})^\alpha + \delta_R^\alpha \tag{5.2}$$

where δ_S and δ_R are MPDs of mobile S and R respectively. In (5.2), it is assumed that the mobile stations and the base station have the same transmitter and receive pair, and that the minimum received signal power required at the front end of a receiver is same for both the mobile stations and the base station. In Figure

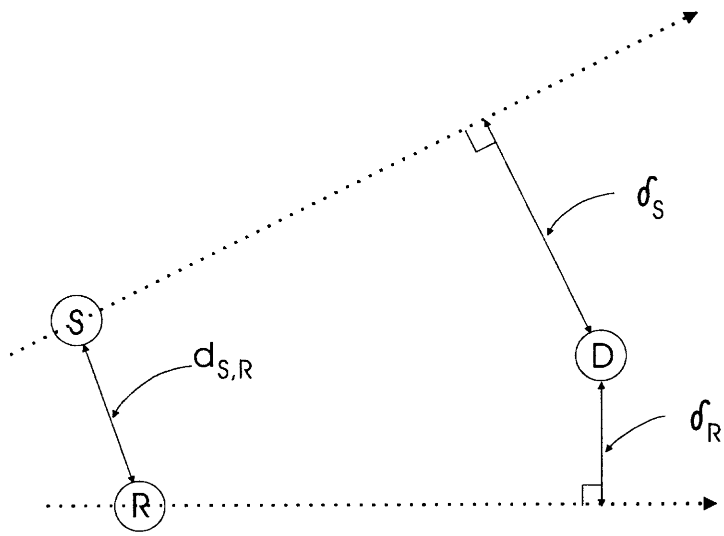


Figure 5.2 Mobility-based multihop routing.

5.2, the trajectory of each mobile station is assumed to be a straight line. If the direction of a mobile station changes, then MPD will also change. Since a mobile makes routing decision based on its MPD and the MPDs of neighboring mobiles, the frequent changes in the direction will result in a ping-pong effect that a same packet is tossed back and forth among mobile stations. Therefore, the movement of relay mobiles should be highly biased in direction. In the mobility-based multihop routing, a small number of highly directional relay mobiles significantly reduce the energy consumption of all other mobiles in the network, that is, a small number of relay mobiles collect user packets from other mobiles and deliver these packets to a base station when they are close to the base station.

Through simulation study, the performance of the mobility-based multihop routing, the minimum energy routing [55] and the conventional single-hop routing is compared with each other. The simulation result shows that the proposed mobility-based multihop routing outperforms the others under reasonable mobility assumptions.

5.2 Brownian Mobility Model

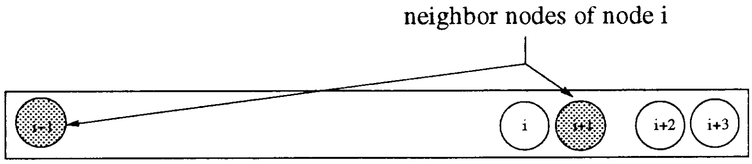
In the simulation study, the Brownian motion with drift (i.e., $B(\mu, \sigma)$ [43]) is used to model individual user mobility.

If variance parameter σ^2 is zero, the mobile moves with constant velocity, and there is no uncertainty about the future position of the mobile. One can predict the position of the mobile at anytime using the current position and the drift velocity. For those whose variance parameter is not zero, a simple trajectory estimation based on the position information and the instantaneous velocity is assumed to determine the minimum power distance of mobiles. Assuming that each mobile is capable of determining its current position, mobiles can determine their instantaneous velocity based on the recent movement history (i.e., the displacement during the last T units of time; the velocity monitoring interval, T, is assumed to be a system parameter). Assuming that all the mobiles in the cell know the coordinates of the base station, each mobile estimates its MPD using its instantaneous velocity. Since the instantaneous velocity is recalculated every T units of time, MPD will also change every T.

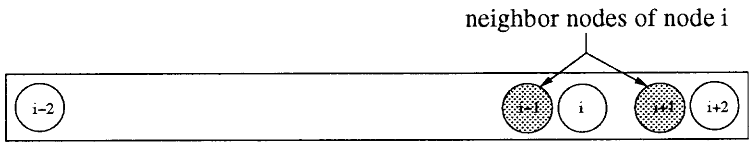
5.3 Energy Efficient Multihop Routing

In this section, three multihop routing algorithms are proposed namely, (*a*) Distance-based Multihop Routing for fixed multihop cellular networks, (*b*) Distance-based Multihop Routing with limited scope of search for next hop node [2] for the fixed multihop cellular network, and (*c*) Mobility-based Multihop Routing which takes advantage of user mobility to reduce energy consumption in the multihop cellular network. The distance-based multihop routing algorithms (*a*) and (*b*) will serve as a basic framework for the mobility-based multihop routing (*c*). Before presenting the

[2] In case where all mobile stations are stationary, the term node is used instead of the term mobile.



(a) Network 1



(b) Network 2

Figure 5.3 Minimum energy routing.

three routing algorithms, the concept of minimum energy routing [55] is described for performance benchmarking purposes. As the term "minimum energy" indicates, the minimum energy routing is to route user data from a source node to a destination node along a multihop route that consumes the least amount of energy. The main challenge in the minimum energy routing is to find the minimum energy route among all possible multihop routes. A distributed routing algorithm which provides the minimum energy routes in the multihop wireless network is proposed in [55]. In the distributed routing algorithm, each node i searches for a minimum set of neighboring nodes $\aleph(i)$ that includes the first hop nodes of all the minimum energy routes from node i to other nodes in the network. In other words, the first hop radio link on a minimum energy route that originates from node i should be one of the radio links between node i and neighboring nodes in $\aleph(i)$. To demonstrate the complexity of the neighbor search, consider the one-dimensional multihop wireless networks in Fig 5.3. If node i examines closer nodes first, the search sequence for node i in Network 1 is $i+1$, $i+2$, $i+3$, $i-1$. Since the neighboring nodes of i (i.e., $\{k : k \in \aleph(i)\}$) are node $i-1$ and node $i+1$, thorough search is required to find all the neighbors of node

i. In Network 2, the search sequence of node i is $i-1$, $i+1$, $i+2$, $i-2$; however, the search can be terminated at $i+1$ since node i already find all the neighboring nodes (the maximum number of neighbors in one-dimensional network is two). In a slightly different search scenario, if node i examines the closest node first (the closest node is the first neighbor to be found) and searches in the opposite direction of the first found neighbor, then the search will terminate in two steps for both networks in Figure 5.3. The search sequence in Network 1 will be $i+1$, $i-1$, and the search sequence in Network 2 will be $i-1$, $i+1$. Constructing efficient search sequence based on the information about already found neighbors in the two-dimensional network is still an open problem. After each node determines its neighboring nodes, each node constructs a link cost table for the radio links between itself and its neighboring nodes. In the distributed minimum energy routing [55], a shortest path algorithm is used to distribute the link cost information of the network where the network is viewed as a sparse graph whose vertex set is ψ and whose edge set is

$$\bigcup_{i \in \psi} \bigcup_{k \in \aleph(i)} l_{i \to k} \tag{5.3}$$

where ψ is the set of all nodes in the network, and $l_{i \to k}$ is the directed radio communication link from node i to node k. When the nodes in the network are fixed, the minimum energy routing is an optimal solution that minimizes energy consumption; however, as the nodes start moving, the overhead for finding the optimal route increases dramatically. For example, if node i changes its position, then the corresponding link cost will change accordingly. The changes in the link cost will prompt a new search of neighboring nodes for node i and the other nodes around node i. Since each node determines the minimum energy route based on the link cost table of other nodes in the network, the minimum energy consumption is not guaranteed unless the link cost is updated in real time. In the simulation study, our implementation of the minimum energy routing [55] assumes that the link costs are updated in

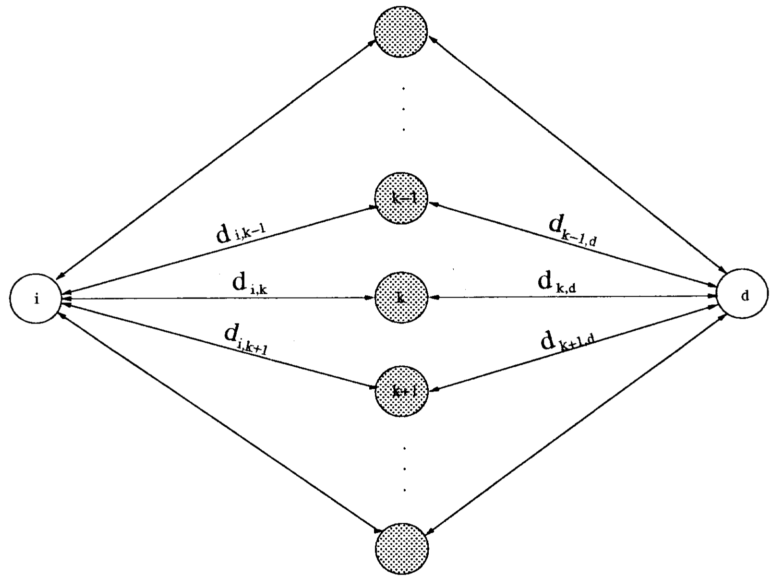


Figure 5.4 Distance-based multihop routing.

real time without any signaling cost; therefore, the "minimum energy consumption" is guaranteed even when nodes are moving.

5.3.1 Distance-based Multihop Routing

A simple distributed multihop routing algorithm is illustrated in Figure 5.4. When node i has a packet to send to destination node d, node i relays its packet to node k^* if there exists node k^* such that

$$\begin{aligned} (d_{i,d})^\alpha &> (d_{i,k^*})^\alpha + (d_{k^*,d})^\alpha \\ k^* &= arg \min_{k\in\psi}\{(d_{i,k})^\alpha + (d_{k,d})^\alpha\}. \end{aligned} \tag{5.4}$$

If node i can't find node k^* that satisfies (5.4) in the network, node i directly sends the packet to destination node d. Node i tries to find a two-hop route to destination node d (i.e., $i \to k^* \to d$), which consumes the least amount of energy among all possible two hop routes. When node k^* receives the packet from node i, k^* will look for its own two-hop route to the destination, which consumes the least amount of energy among all possible two hop routes. The next relay node that receives the packet from node k^* will do the same. Therefore, the total hop count from node i to node d can be greater than two. When node i makes routing decision, node i needs to know the cost (i.e., the power requirement) of the radio link between itself and the potential relay nodes and the link between the potential relay nodes and the destination node. If the distance between node i and node k is larger than the distance between node i and the destination, then node k cannot satisfy (5.4), and it shouldn't be considered as a relay. Therefore, only those nodes that are closer than the destination are considered as potential relay nodes. Assuming that each node is capable of self-geolocation and that they share the position information with other nodes in the network to determine link cost, node i needs the position information of other nodes that are closer than the destination node. Sharing position information with other nodes in the network will require significant signaling overhead. In the following, the scope of search for the next hop node is limited to increase the scalability of the algorithm.

5.3.2 Distance-based Multihop Routing with Limited Scope of Search for Next Hop Node

Comparing to the minimum energy routing [55] in which each node requires the knowledge of the network topology (i.e., the sparse graph and the link cost) to determine the minimum energy route, the distance-based multihop routing is much more scalable since only the nodes that are closer than the destination node are considered as potential relay nodes. Node i searches for a next hop node in a circular area of radius $d_{i,d}$ centered at node i where $d_{i,d}$ is the distance between node i and destination node d. By reducing the size of the search area, the scalability of the algorithm increases. The radius of the search area is referred as Power Saving Range (PSR).

Definition 1 *The nodes that are in the power saving range of node i are neighbors of node i, and the set of neighbors of node i is denoted as $\aleph'(i)$.*

Notice that the definition of neighbor in the minimum energy routing [55] differs from Definition 1. When node i has a packet to send to destination node d, node i relays the packet to node k^* if there exists node k^* such that

$$\begin{aligned} (d_{i,d})^\alpha &> (d_{i,k^*})^\alpha + (d_{k^*,d})^\alpha \qquad (5.5) \\ k^* &= arg \min_{k \in \aleph'(i)} \{(d_{i,k})^\alpha + (d_{k,d})^\alpha\}. \end{aligned}$$

If node i cannot find neighboring node k^* that satisfies (5.5), node i directly transmits the packets to destination node d.

5.3.3 Mobility-based Multihop Routing

As mentioned earlier, relay mobiles in the mobility-based multihop routing should be highly biased in direction to avoid the ping-pong effect. If the relay mobiles change their movement directions frequently, the MPDs of the mobiles will also change frequently, and the mobiles will toss a same packet back and forth wasting transmission energy. Therefore, a relay mobile should have a large drift velocity and a small variance parameter to have highly directional movement pattern. Following two conditions are imposed for relay mobiles:

$$|\mu| > K_1, \quad \sigma^2 < K_2 \qquad (5.6)$$

where

- K_1: minimum drift velocity for relay mobiles,
- K_2: maximum variance parameter for relay mobiles.

The definition of a relay mobile in the mobility-based multihop routing is defined as follows.

Definition 2 *A mobile which satisfies (5.6) is a relay mobile. Otherwise, it is a non-relay mobile.*

Both the relay and the non-relay mobiles are traffic sources that can generate data packets destined to base stations. If K_1 is large and K_2 is small, then only the mobiles that are highly biased in the direction are allowed to become relay mobiles. The minimum power distance to base station (MPD) of a non-relay mobile is assumed to be infinite. By setting the MPDs of the non-relay mobiles to be infinite, mobiles will not choose a non-relay mobiles as their next hop. By avoiding packet relay to a mobile that changes direction frequently, the ping-pong effect can be reduced significantly.

The MPD of a relay mobile which is moving away from its current base station is also considered to be infinite. In the mobility-based multihop routing, a mobile does not directly transmit data packets to a base station if the distance between the mobile and the base station is greater than PSR. This is an artificially imposed radio communication range. The idea is to avoid long range direct radio transmission from mobiles to base stations. By artificially limiting the radio communication range, the power consumption will be reduced.

It is assumed that two neighboring mobiles share their position information and their MPDs, and each mobile knows the position of its current base station. If mobile i has a packet destined for a base station, mobile i searches for a next hop mobile among its neighbors using one of the following routing rules depending on its current position.

5.3.3.1 Routing Rule 1. When the distance from mobile i to its base station is greater than PSR, mobile i relay its packet to mobile k^* if there exists mobile k^*

such that

$$\begin{aligned} \delta_i^\alpha &> (d_{i,k^*})^\alpha + \delta_{k^*}^\alpha \\ k^* &= arg \min_{k \in \aleph'(i)} \{(d_{i,k})^\alpha + \delta_k^\alpha\} \end{aligned} \tag{5.7}$$

where $d_{i,k}$ is the distance between mobile i and mobile k, and δ_i and δ_k are the MPDs of mobile i and k respectively. If mobile i doesn't have any neighboring mobile satisfying (5.7), the mobile waits until it encounters a mobile that satisfies (5.7).

5.3.3.2 Routing Rule 2. When the distance between mobile i and the base station is less than or equal to PSR, mobile i relay its packet to mobile k^* if there exists mobile k^* such that

$$\begin{aligned} \delta_i^\alpha &> (d_{i,k^*})^\alpha + \beta_{k^*}^\alpha \\ k^* &= arg \min_{k \in \aleph'(i)} \{(d_{i,k})^\alpha + \beta_k^\alpha\} \\ \beta_k &= min(\delta_k, d_{k,d}) \end{aligned} \tag{5.8}$$

where $d_{k,d}$ is the distance between mobile k and the base station. If mobile i doesn't have any neighboring mobile satisfying (5.8), then mobile i makes routing decision based on its MPD. If the MPD of mobile i is finite, mobile i waits until it encounters a mobile that satisfies (5.8). If the MPD is infinite, then mobile i directly transmits its packet to the base station.

5.4 Simulation Model

In the simulation study, the energy consumption of the distance-based multihop routing, the mobility-based multihop routing, the minimum energy routing [55], and the conventional single-hop routing is compared. A simple path-loss propagation model [52] is used to determine the power consumption. The energy spent at a receiver for signal processing as well as the effect of interference due to the multiple

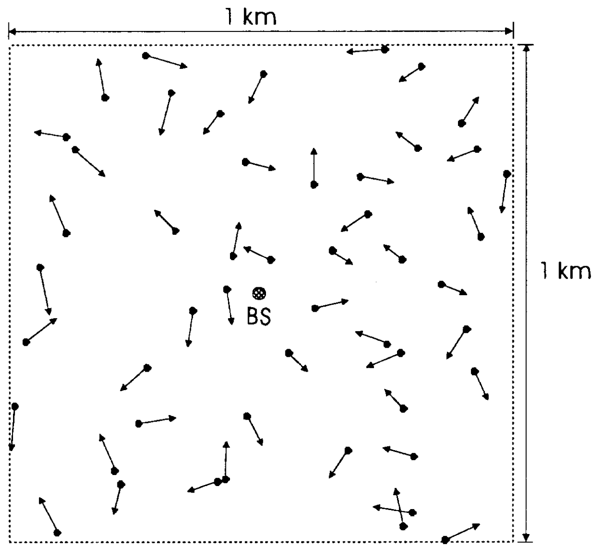


Figure 5.5 Random direction mobility model.

access are not considered, that is, it is assumed that radio transmissions of different mobiles do not occur at the same time, and these radio transmissions do not collide. It is assumed that the minimum received signal power, P, required at the front end of a receiver for reliable packet reception is same for all mobile stations and base stations regardless of the routing protocol employed. The required transmission power at a transmitter to reach a receiver which is d km apart is Pd^{α}/G_tG_r where G_t and G_r are the transmitter and the receiver antenna gains [52]. The path loss exponent, α, is assumed to be 4. The same transmitter and receiver pair is assumed for mobile stations and base stations. The energy required to deliver a packet from a mobile to a base station is the sum of the required transmission power on each radio link along the route to the base station multiplied by the packet transmission delay τ. The packet transmission delay is assumed to be same for all packets. For simplicity, the energy consumption is normalized by $P\tau/G_tG_r$.

While the movement of an individual mobile is modeled using the Brownian motion, two cell-level mobility models, namely, random direction mobility model and grid mobility model (see Figure 5.5 and 5.6 respectively) is used to simulate

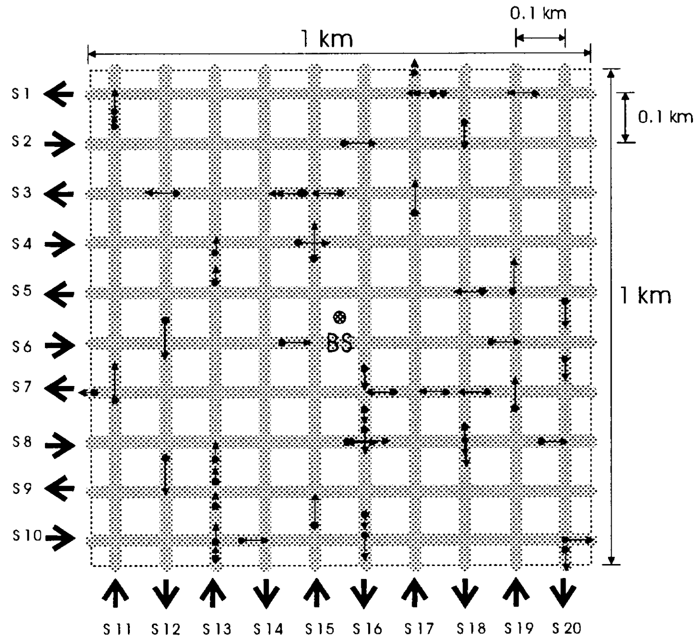


Figure 5.6 Grid mobility model.

a single cell. The shape of the cell is square, and the area of the cell is 1km $\times$ 1km. In the random direction model, the initial position of a mobile is uniformly distributed over the cell. For each mobile, the angle of drift velocity (i.e., $\angle\mu$) is uniformly distributed between 0 and 2π. The same amplitude of drift velocity (i.e., $|\mu|$) and the same variance parameter are assumed for all mobiles in the cell. The drift velocity and the variance parameter remain constant during the simulation.

The grid mobility model in Figure 5.6 consists of a number of one-way streets and square shaped building blocks. The position of the base station is assumed to be the origin in the Cartesian coordinate system. The initial positions of mobiles are drawn from uniform distribution; however, each mobile is relocated to a nearest street, and the mobiles are only allowed to move along the streets. After the relocation, the mobiles are no longer uniformly distributed over the cell. For example, if the initial position drawn from uniform distribution is (0.04 km, 0.03 km), the closest street from the mobile is $S16$. The coordinate of street $S16$ is $x = 0.05$ km. The mobile is relocated to a closest location on $S16$, which is (0.05 km, 0.03 km). The movement of each mobile is one-dimensional Brownian motion, and mobiles are not allowed to change streets at the intersections. The same amplitude of drift velocity (i.e., $|\mu|$) and the same variance parameter are assumed for all mobiles in the cell.

For both the random direction model and the grid model, the mobiles that reach the cell boundary reenter the cell from the opposite side of the boundary; therefore, the number of mobiles in the cell remains constant during the simulation. Each mobile generates packets destined to the base station according to a Poisson arrival process with rate λ packets/min. Each mobile has a FIFO queue with infinite buffer space. A point-to-point radio link is modeled as a bit pipe where there is no interference from other radio links. The radio propagation delay is assumed zero. There is no restriction on the number of hops that a packet can travel before it reaches the base station.

For the mobility-based multihop routing, the radio communication range is artificially limited to PSR. It is assumed that each mobile is aware of the position coordinates of the base station and the position coordinates and MPDs of neighboring mobiles within the PSR.

For the minimum energy routing [55], it is assumed that each mobile is aware of the position coordinates of the base station and the other mobiles in the cell without any additional cost. Using the position coordinates of mobiles, each mobile determines the minimum energy multihop route from itself to the base station. The base station can be considered as a fixed node that is the destination of all the mobiles in the cell. Among all possible multihop routes, a mobile chooses a route that minimizes the sum of the transmission energy required on each point-to-point radio link along the route. There is no restrictions on the communication range as opposed to the mobility-based multihop routing. In fact, limiting the communication range in the minimum energy routing will increase the power consumption. If a mobile

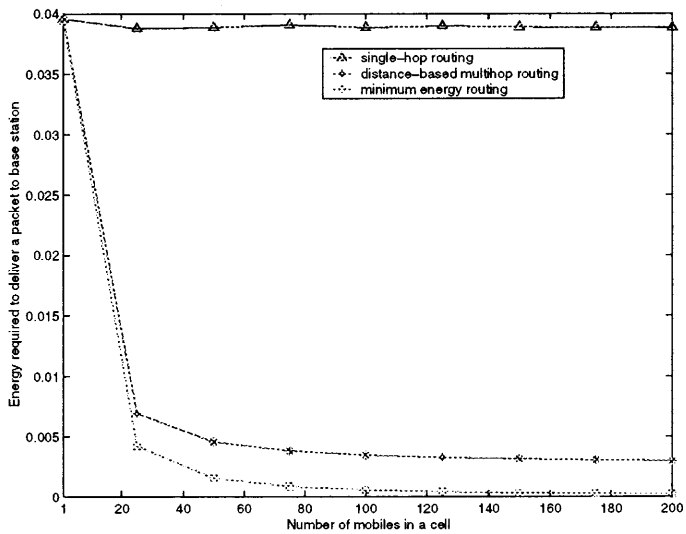


Figure 5.7 Energy consumption vs. the number of mobiles in a cell for the random direction mobility model with $|\mu| = 0$, $\sigma^2 = 0$, $\alpha = 4$, $\lambda = 10$, and $N = 1, 25, 50, 75, 100, 125, 150, 175, 200$.

cannot choose a route that minimizes the energy consumption due to the limited communication range, the mobile will choose an alternate route that consumes more energy than the minimum energy route.

5.5 Simulation Results

5.5.1 Random Direction Mobility Model

First, the energy consumption of the fixed wireless network is simulated using the random direction mobility model where the drift velocity and the variance parameter are zeroes for all mobiles. The positions of mobiles are uniformly distributed. Figure 5.7 shows the energy consumption of the three routing algorithms: the distance-based multihop routing without limited scope of search for next hop node, the minimum energy routing, and the conventional single-hop routing. As the number of mobiles in the cell, N, increases, the energy consumption of the distance-based multihop routing and the minimum energy routing decreases rapidly. For $N = 1$, there is no

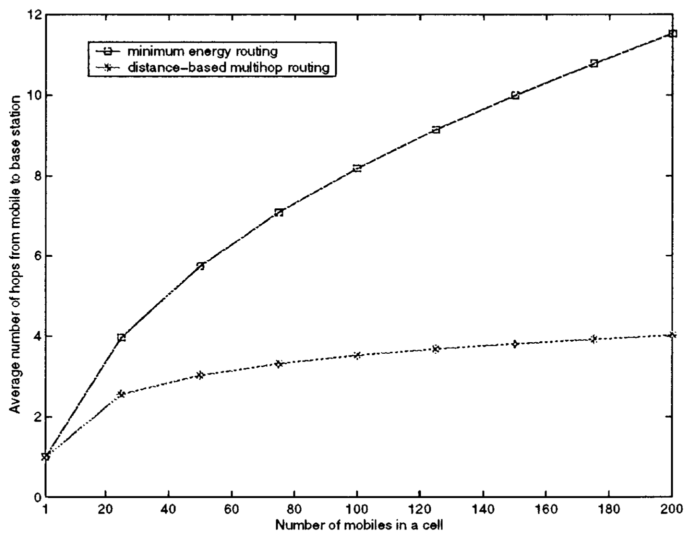


Figure 5.8 Average number of hops vs. the number of mobiles in a cell for the random direction mobility model with $|\mu| = 0$, $\sigma^2 = 0$ $\alpha = 4$, $\lambda = 10$, and $N = 1, 25, 50, 75, 100, 125, 150, 175, 200$.

difference in the performance of the three routing algorithms. The mobile transmits packets directly to the base station. For $N \geq 25$, the energy consumption of the minimum energy routing and the distance-based routing saturates. At $N = 25$, the minimum energy routing consumes only 10.9% of the energy required by the conventional single-hop routing while the distance-based routing consumes 17.9% of the energy required by the single-hop routing. The minimum network density needed for meaningful energy saving (i.e., more than 80% of reduction in the energy consumption) is about 25 mobiles/km^2 assuming that the path loss exponent is $\alpha = 4$. Figure 5.8 shows the average number of hops that a packet travels until it arrives at the base station. For the single-hop routing, the hop count is always 1. As N increases, the average hop count for the minimum energy routing increases much faster than the average hop count for the distance-based routing. For $N = 75$, the hop count for the minimum energy routing doubles the hop count for the distance-based routing. The ratio between the hop counts for the two multihop routing algorithms keeps increasing as N increases. Figure 5.9 shows the energy

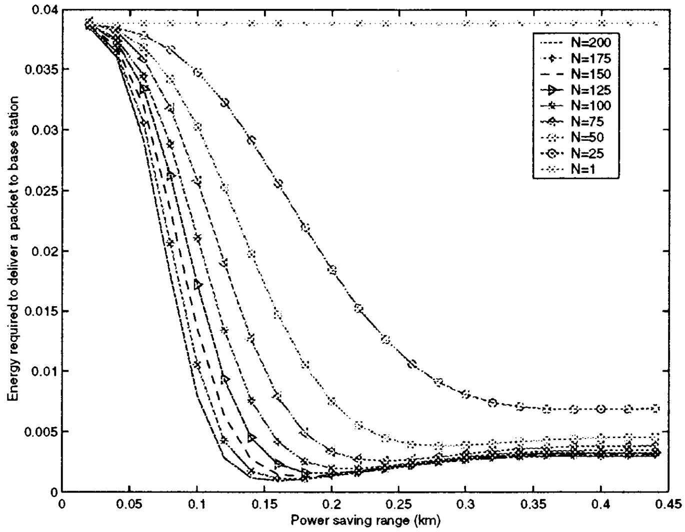


Figure 5.9 Energy consumption vs. power saving range (PSR) in the distance-based multihop routing with limited scope of search for the random direction mobility model with $|\mu| = 0$, $\sigma^2 = 0$, $\lambda = 10$, and $N = 1, 25, 50, 75, 100, 125, 150, 175, 200$.

consumption of the distance-based multihop routing with limited scope of search. The x-axis is the power saving range (PSR) that is the radius of the search area. Except for $N = 1$, there exists local minimum for each curve. In the distance-based multihop routing, mobiles search for a next hop mobile that satisfies (5.5). If PSR is too small, most of the mobiles won't be able to find a next hop mobile, and they will transmit packets directly to the base station. For $N = 200$ and PSR=0.02 km, the energy consumption of the distance-based multihop routing is 0.0384 while the energy consumption of the conventional single-hop routing is 0.0388. As PSR increases, the energy consumption of the distance-based multihop routing decreases rapidly. As PSR increases, mobiles have better chance to find a next hop mobile that satisfies (5.5); therefore, the long range direct radio transmission to the base station is likely avoided. On the other hand, the average per-hop distance increases as PSR increases. This is why the energy consumption increases after the local minimum. For $N = 200$, the energy consumption at PSR=0.16 km is 2.4% of the energy that is required by the single-hop routing. The minimum energy routing consumes 0.5% of the energy

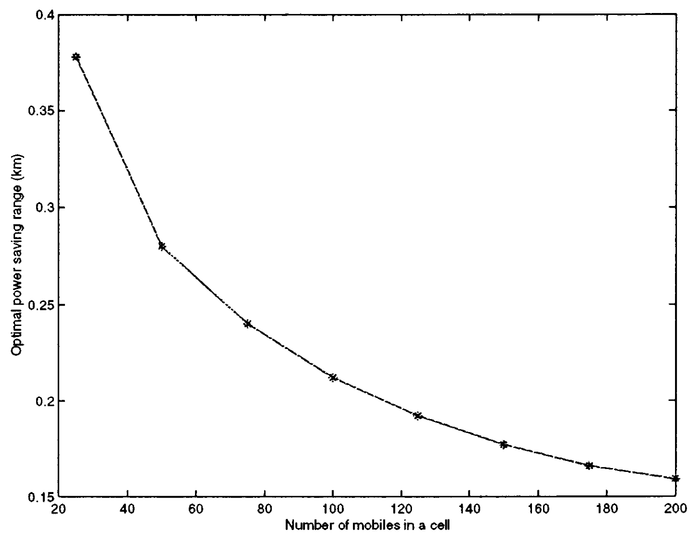


Figure 5.10 Optimal power saving range (PSR) vs. the number of mobiles in the distance-based multihop routing with limited scope of search for the random direction mobility model with $|\mu| = 0$, $\sigma^2 = 0$, $\lambda = 10$, and $N = 25, 50, 75, 100, 125, 150, 175, 200$.

required by the single-hop routing (see Figure 5.7). The larger the number of mobiles in the cell, the higher the probability that a mobile will find an energy efficient next hop mobile with given PSR. That is why the energy consumption decreases more rapidly when N is large. In Figure 5.10, the optimal PSR value that minimizes the energy consumption is plotted. As the number of mobiles increases, the optimal size of the area in which each mobile searches for an energy efficient next hop mobile decreases. Since each mobile does not need to be concerned about other mobiles that are outside of its search area, the proposed multihop routing algorithm is scalable. For $N = 25$, the optimal PSR value is 0.38 km, and the average number of mobiles in the optimal search area is $2\pi \cdot (PRS)^2 \cdot N$=3.6 mobiles. For $N = 200$, the optimal PSR value is 0.16 km, and the average number of mobiles in the optimal search area is $2\pi \cdot (PRS)^2 \cdot N$=5.1 mobiles. While the number of mobiles in the cell increases from N=25 to N=200, the signaling overhead per mobile increases only 40% (i.e., 3.6 mobiles/search area to 5.1 mobiles/search area). While the minimum energy routing

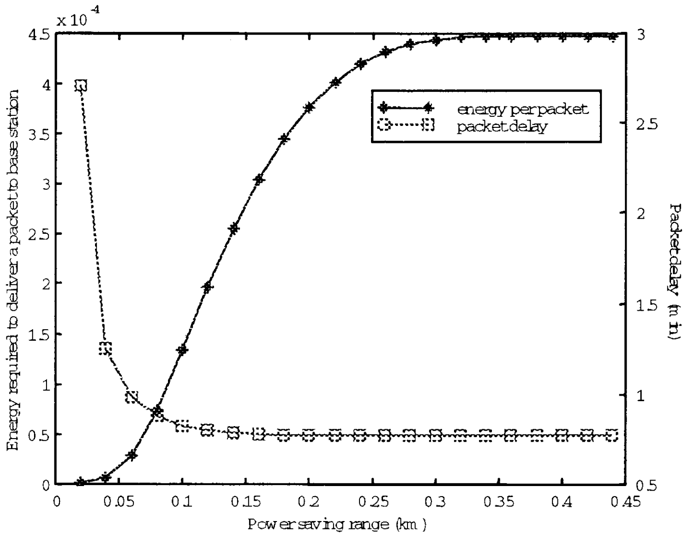


Figure 5.11 Energy consumption and packet delay vs. power saving range (PSR) for the random direction mobility model with $|\mu| = 0.5$, $\sigma = 0.01$, $\lambda = 10$, and $N = 200$.

[55] provides an optimal multihop route that minimizes the energy consumption in the fixed wireless network, the proposed distance-based multihop routing provides a scalable suboptimal solution.

In the following, the performance of the mobility-based multihop routing algorithm is analyzed. Figure 5.11 shows the energy consumption and the packet delay for the random direction mobility model with drift velocity $|\mu| = 0.5$ km and variance parameter $\sigma^2 = 0.01$ km^2/min. It is assumed that there are 200 mobiles in the cell where each mobile generates data packets with rate $\lambda = 10$ packets/min. All mobiles are assumed to be relay mobiles as defined in Definition 2. The mobiles route their packets based on the routing rules (5.7) and (5.8) to take advantage of the user mobility to reduce energy consumption. The energy consumption increases rapidly until PSR=0.3 km. For PSR > 0.3, the energy consumption saturates to 4.47×10^{-4}. The reason for the saturation is that mobiles choose relay mobiles that are close to themselves. A relay mobile that is far away from the transmitting mobile is likely disqualified as a next hop relay mobile since it takes long range

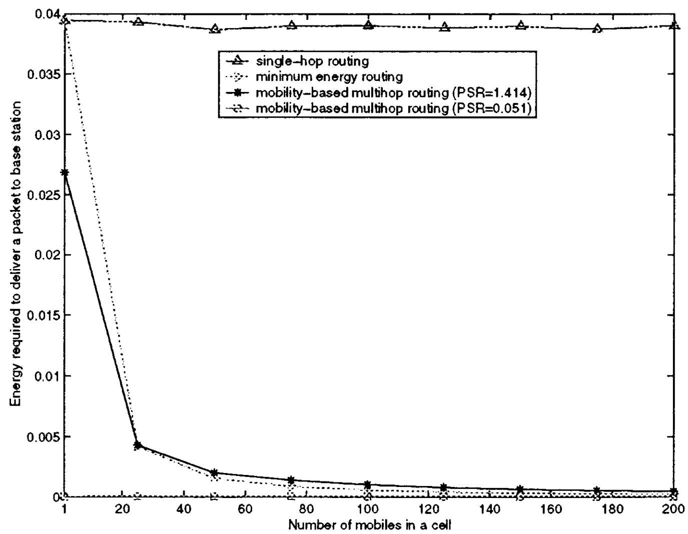


Figure 5.12 Energy consumption vs. the number of mobiles in a cell for the random direction mobility model with $|\mu| = 0.5$, $\sigma^2 = 0.01$, $\lambda = 10$, and $N = 1, 25, 50, 75, 100, 125, 150, 175, 200$.

radio transmission to reach the relay mobile. The simulation shows that the energy consumption decreases as PSR decreases. The trade-off for having small energy consumption is packet delay. In Figure 5.11, the packet delay decreases rapidly as PSR increases until PSR=0.1. For PSR>0.1, the packet delay saturates to 46.2 seconds. PSR should be small to reduce the energy consumption, and at the same time the PSR should be large enough to avoid excessive packet delay. Figure 5.12 shows the energy consumption for different number of mobiles in the cell. The drift velocity, the variance parameter and the packet rate for each mobile are $|\mu| = 0.5$ km/min, $\sigma^2 = 0.01$ km^2/min and $\lambda = 10$ packets/min respectively. For the mobility-based multihop routing, two curves are shown in the figure, that is, one for PSR=∞ and the other for PSR=0.051 km. Based on the previous simulation result, it is expected that the energy consumption will be small if PSR is small. The energy consumption for the conventional single-hop routing remains constant as the number of mobiles increases from N=1 to N=200. Since mobiles directly transmit their packets to the base station, the energy consumption of a

Table 5.1 Energy Consumption of the Mobility-based Multihop Routing with PSR=0.051

Number of mobiles	1	25	50	75	100
Energy ($\times 10^{-5}$)	0.3	2.1	2.0	1.9	1.8

Number of mobiles	125	150	175	200
Energy ($\times 10^{-5}$)	1.7	1.7	1.7	1.7

mobile station is independent of the number of mobiles in the cell. Because mobiles are moving, and their distance to the base station keeps changing, the long term average power consumption per mobile will be same for all mobiles in the cell. For $N = 1$, the single-hop routing and the minimum energy routing consumes the same amount of energy since the mobile directly transmits packets to the base station. In the mobility-based multihop routing, the mobile does not transmit packets to the base station if the distance to the base station is greater than PSR. For N=200 and PSR=0.051, the energy required to deliver a packet to the base station in the mobility-based multihop routing is 1.7×10^{-5} which is 0.044% of the energy required by the conventional single-hop routing, and the packet delay is slightly greater than 1 minute (see Fig 5.11). The packet delay will decrease if the drift velocity increases.

For PSR=0.051, Figure 5.12 doesn't clearly show how the energy consumption of the mobility-based multihop routing changes as N increases. In Table 5.1, the simulation result is tabulated for clarity. The energy consumption increases as the number of mobiles increases from N=1 to N=25. When there is only one mobile in the cell, there is no multihop relay. The mobile directly transmits its packets to the base station if MPD is smaller than PSR, and the distance to the base station becomes MPD. Figure 5.13 shows why the energy consumption increases as the number of mobiles increases from N=1 to N=25. Assume that the distance between mobile i and j is very small such that $(d_{i,j})^\alpha \approx 0$. Also assume that mobile i and j

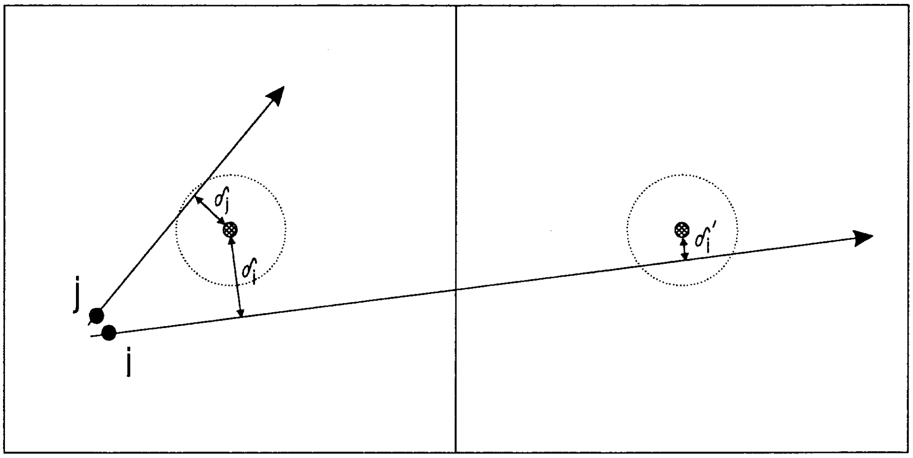


Figure 5.13 Two-hop relay that consumes more energy than one-hop relay.

are moving with constant velocity, and that $\delta_i > \delta_j \geq \delta_i'$. Since $(\delta_i)^\alpha > (d_{i,j})^\alpha + (\delta_j)^\alpha$, mobile i will relay its packet to mobile j, and the energy required to deliver the packet to the base station is $(d_{i,j})^\alpha + (\delta_j)^\alpha$. If mobile i didn't relay its packet and crossed the cell boundary, mobile i would deliver the packets to the base station in the next cell. The energy required to deliver the packet to the neighboring base station is $(\delta_i')^\alpha$, and that is smaller than $(d_{i,j})^\alpha + (\delta_j)^\alpha$ that is the energy required to deliver the packet to the current base station using multihop relay. Therefore, the energy consumption would be smaller if mobiles do not relay packets. For $N > 25$, the energy consumption decreases as N increases. It is because each mobile gets better chance to find a "good" relay mobile in its neighborhood as the number of mobiles in the cell increases.

For PSR=∞, the energy consumption of the mobility-based multihop routing is close to that of the minimum energy routing except for $N = 1$. For $N = 200$, the minimum energy routing consumes 0.5% of the energy required by the conventional single-hop routing while the mobility-based multihop routing consumes 1.2% of the energy required by the single-hop routing.

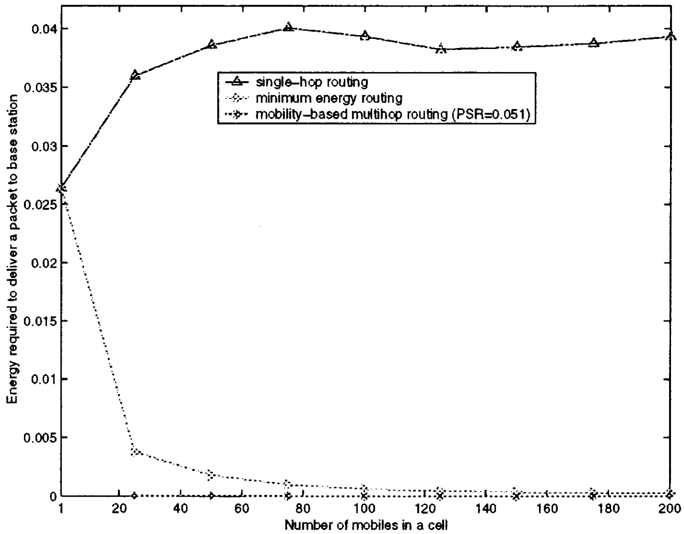


Figure 5.14 Energy consumption vs. the number of mobiles in a cell for the grid mobility model with $|\mu| = 0.5$, $\sigma^2 = 0.01$, $\lambda = 10$, and $N = 1, 25, 50, 75, 100, 125, 150, 175, 200$.

5.5.2 Grid Mobility Model

For $|\mu| = 0.5$ km and $\sigma^2 = 0.01$ km^2/min, the average energy consumption for the grid mobility model is plotted in Figure 5.14. As in the random direction model, the energy consumption of the minimum energy routing is same as the energy consumption of the single-hop routing for $N = 1$. In our simulation for $N = 1$, the mobile happens to land on street S18 (see Figure 5.6). The minimum distance to the base station for a mobile on Street S18 is relatively smaller comparing to mobiles on the other streets; therefore, the energy consumption for $N = 1$ is smaller than the average energy consumption for large N.

For $N = 200$ and PSR=0.051 km, the energy consumption of the mobility-based multihop routing is 0.038% (c.f., 0.044% for random direction mobility model) of the energy required by the conventional single-hop routing while the energy consumption of the minimum energy routing is 0.63% (c.f., 0.5% for random direction mobility model) of the energy required by the single-hop routing.

As N increases, the energy consumption of the single-hop routing stabilizes to 3.93×10^{-2} for the grid mobility model while the energy consumption of the single-hop routing stabilizes to 3.86×10^{-2} for the random direction mobility model (see Figure 5.12).

The performance of the minimum energy routing and the mobility-based multihop routing becomes close to that of the random direction mobility model as the number of mobile increases.

CHAPTER 6
CONCLUSION

This dissertation presents an predictability-based approach to mobility modeling and management in wireless networks. As opposed to the conventional mobility modeling approach that considers the speed of mobile as the level of mobility such as "high" or "low" mobility, the predictability-based approach defines mobility as the uncertainty in mobile's motion pattern. A high speed mobile may have *low* mobility if the uncertainty in its motion pattern is small, and the network can predict the trajectory of the mobile accurately. On the other hand, a slow mobile could have *high* mobility if its trajectory is highly erratic, that is, the uncertainty in its motion pattern high.

Based on this new mobility modeling concept, trajectory prediction is proposed to reduce the cost of mobility related transactions in wireless networks. In the location management, a mobile on a highly directed path such as highway or railroad does not need to update its location frequently even if the mobile is moving with high speed since the network can predict the location of the mobile when a call or service arrives. In the location-based services, the trajectory prediction reduces the cost of geolocation updates for those mobiles that are highly biased in direction. In the multihop cellular network, mobiles that do not change their direction frequently may serve as relays and deliver data traffic from other mobiles to cellular base stations in an energy efficient manner.

Analysis based on a Brownian motion model shows that the trajectory prediction is not always feasible. For those mobiles whose trajectory is *erratic*, the trajectory prediction actually increases the cost of the mobility related transactions. The feasibility region of the trajectory prediction is determined through analysis, and there exists an optimal mobility classification rule that classifies mobility patterns into *predictable* and *non-predictable* mobility classes. As far as the network operation cost is concerned, the network should not employ trajectory prediction if a mobile belongs to *non-predictable* mobility class, and the network should employ trajectory prediction if a mobile belongs to *predictable* mobility class.

The predictability-based mobility modeling approach and the dynamic mobility classification technique can be beneficial to virtually every other mobility management issues that are not treated in this dissertation.

REFERENCES

1. S. Tekinay and B. Jabbari, "Handover and Channel Assignment in Mobile Cellular Networks," *IEEE Communications Magazine*, vol. 29, no. 11, pp. 42–46, Nov. 1991.

2. R. A. Guerin, "Channel Occupancy Time Distribution in a Cellular Radio System," *IEEE Transactions on Vehicular Technology*, vol. 35, no. 3, pp. 89–99, Aug. 1987.

3. J. Broch , D. A. Maltz, D. B. Johnson, Y.-C. Hu and J. Jetcheva , "A Performance Comparison of Multi-hop Wireless Ad Hoc Network Routing Protocols," in *Proc. ACM International Conference on Mobile Computing and Networking*, Oct. 1998, pp. 85–97.

4. D. A. Levine, I. F. Akyildiz and M. Naghshineh, "A Resource Estimation and Call Admission Algorithm for Wireless Multimedia Networks using the Shadow Cluster Concept," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 1–12, Feb. 1997.

5. D. Gu and S. S. Rappaport, "A Dynamic Location Tracking Strategy for Mobile Communication Systems," in *Proc. IEEE Vehicular Technology Conference Fall*, 1998, pp. 259–263.

6. A. K. Talukdar, B. R. Badrinath and A. Acharya, "On Accommodating Mobile Hosts in an Integrated Services Packet Network," in *Proc. IEEE INFOCOM Conference*, Apr. 1997, pp. 1046–1053.

7. S. Tekinay, *Modeling and Analysis of Mobile Cellular Networks with Highly Mobile Heterogeneous Traffic Sources*, Ph.D. thesis, George Mason University, Virginia, USA, 1994.

8. C. Rose and R. Yates, "Location Uncertainty in Mobile Networks: A Theoretical Framework," *IEEE Communications Magazine*, vol. 35, no. 2, pp. 94–101, Feb. 1997.

9. C. Rose and R. Yates, "Minimizing the Average Cost of Paging Under Delay Constraints," *ACM Wireless Networks*, vol. 1, no. 2, pp. 211–219, 1995.

10. W. Choi, and S. Tekinay, "An Adaptive Location Registration Scheme with Dynamic Mobility Classification," in *Proc. IEEE International Conference on Communications*, 2002, pp. 440–444.

11. H. D. Robertson, J. E. Hummer and D. C. Nelson, *Manual of Transportation Engineering Studies*, Institute of Transportation Engineers, Prentice Hall, NJ, 1994.

12. E. Aleman-Llanes, D. Munoz-Rodriguez and C. Molina, "FBM Mobility Modeling for Nomadic Subscribers," in *Proc. IEEE Symposium on Computers and Communications*, Athens, Greece, Jun. 1998, pp. 315–319.

13. R. Thomas, H. Gilbert and G. Maziotto, "Influence of the Moving of the Mobile Stations on the Performance of a Radio Mobile Cellular Network," in *Proc. the 3rd Nordic Seminar on Digital Land Mobile Radio Communications*, Sep. 1988.

14. H. Xie, S. Tabbane and D. J. Goodman, "Dynamic Location Area Management and Performance Analysis," in *Proc. IEEE Vehicular Technology Conference*, May 1993, pp. 536–539.

15. W. F. Fuhrmann and V. Brass, "Performance Aspects of a GSM Radio Subsystem," *Proceedings of the IEEE*, vol. 82, no. 9, pp. 1449–1466, Sep. 1994.

16. D. Lam, D. C. Cox and J. Widom, "Teletraffic Modeling for Personal Communications Services," *IEEE Communications Magazine*, vol. 35, no. 2, pp. 79–87, Feb. 1997.

17. M. Sidi and D. Starobinski, "New Call Blocking versus Handoff Blocking in Cellular Networks," *ACM Wireless Networks*, vol. 3, no. 1, Mar. 1997.

18. K. Okada and F. Kubota, "On dynamic channel assignment in cellular mobile radio systems," in *Proc. IEEE International Symposium on Circuits and Systems*, Jun. 1991, pp. 938–941.

19. C. Chao and W. Chen, "Connection Admission Control for Mobile Multiple-Class Personal Communications Networks," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 8, pp. 1618–1626, Oct. 1997.

20. H. Kobayashi, S. Yu and B. L. Mark, "An Integrated Mobility and Traffic Model for Resource Allocation in Wireless Networks," in *Proc. ACM International Workshop on Wireless Mobile Multimedia*, Boston, MA, Aug. 2000, pp. 39–47.

21. D. Hong and S. S. Rappaport, "Traffic Model and Performance Analysis for Cellular Mobile Radio Telephone Systems with Prioritized and Nonprioritized Handoff Procedures," *IEEE Transactions on Vehicular Technology*, vol. 35, no. 3, pp. 77–92, Aug. 1986.

22. S. A. El-Dolil, W.-C. Wong and R. Steele, "Teletraffic Performance of Highway Microcells with Overlay Macrocell," *IEEE Journal on Selected Areas in Communications*, vol. 7, no. 1, pp. 71–78, Jan. 1989.

23. D. E. Everitt, "Traffic Engineering of the Radio Interface for Cellular Mobile Networks," *Proceedings of the IEEE*, vol. 82, no. 9, pp. 1371–1382, Sep. 1994.

24. E.-S. El-Alfy, Y.-D. Yao and H. Heffes, "A Learning Approach for Call Admission Control with Prioritized Handoff in Mobile Multimedia Networks," in *Proc. IEEE Vehicular Technology Conference Spring*, 2001, pp. 972–976.

25. I. Rubin and C. W. Choi, "Impact of the Location Area Structure on the Performance of Signaling Channels in wireless Cellular Networks," *IEEE Communications Magazine*, vol. 35, no. 2, pp. 108–115, Feb. 1997.

26. H. Xie and D. J. Goodman, "Mobility Models and Biased Sampling Problem," in *Proc. IEEE International Conference on Universal Personal Communications*, Oct. 1993, pp. 803–807.

27. H. Xie and S. Kuek, "Priority Handoff Analysis," in *Proc. IEEE Vehicular Technology Conference*, May 1993, pp. 855–858.

28. M. M. Zonoozi and P. Dassanayake, "User Mobility Modeling and Characterization of Mobility Patterns," *IEEE Journal on Selected Areas in Communications*, vol. 15, no. 7, pp. 1239–1252, Sep. 1997.

29. P. Orlik and S. S. Rappaport, "A Model for Teletraffic Performance and Channel Holding Time Characterization in Wireless Cellular Communication," in *Proc. IEEE International Conference on Universal Personal Communications*, San Diego, CA, Oct. 1997, pp. 671–675.

30. Y. Fang and I. Chlamtac, "Teletraffic Analysis and Mobility Modeling of PCS Networks," *IEEE Transactions on Communications*, vol. 47, no. 7, pp. 1062–1072, Jul. 1999.

31. S. Tekinay, "Adaptive Networks for Next Generation Wireless Communications: The Growing Role of Peer-to-Peer Communications," submitted for publication.

32. E. D. Kaplan, Ed., *Understanding GPS: Principles and Applications*, Artech House, Boston, MA, 1996.

33. K. Pahlavan, Li Xinrong, J.P. Makela, "Indoor Geolocation Science and Technology," *IEEE Communications Magazine*, vol. 40, no. 2, pp. 112–118, Feb. 2002.

34. W. Choi, and S. Tekinay, "An Efficient Table Driven Routing Algorithm for Wireless Ad Hoc Networks," in *Proc. IEEE Vehicular Technology Conference Fall*, 2001, pp. 2604–2608.

35. S. Basagni, I. Chlamtac and V. R. Syrotiuk, "A Distance Routing Effect Algorithm for Mobility (DREAM)," in *Proc. ACM International Conference on Mobile Computing and Networking*, 1998, pp. 76–84.

36. B. Liang and Z. Haas, "Predictive Distance-based Mobility Management for PCS Networks," in *Proc. IEEE INFOCOM Conference*, 1999, pp. 1377–1384.

37. A. Aljadhai and T. F. Znati, "Predictive Mobility Support for QoS Provisioning in Mobile Wireless Environments," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 10, pp. 1915–1930, 2001.

38. U. Madhow, M. Honig and K. Steiglitz, "Optimization of Wireless Resources for Personal Communications Mobility Tracking," *IEEE/ACM Transactions on Networking*, vol. 3, no. 4, pp. 698–707, 1995.

39. J. Ho and I. F. Akyildiz, "Mobile User Location Update and Paging Under Delay Constraints," *ACM Wireless Networks*, vol. 1, no. 4, pp. 413–425, 1995.

40. A. Bar-Noy, I. Kessler and M. Sidi, "Mobile Users: To update or not to update?," *ACM Wireless Networks*, vol. 1, no. 2, pp. 175–185, 1995.

41. S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, Academic Press, San Diego, CA, 2nd edition, 1975.

42. T. S. Rappaport, J. H. Reed and B. D. Woerner, "Position Location Using Wireless Communications on Highways of the Future," *IEEE Communications Magazine*, pp. 33–41, Oct. 1996.

43. J. M. Harrison, *Brownian Motion and Stochastic Flow System*, John Wiley and Sons, New York, NY, 1985.

44. C. Yin, "The Joint Distribution of the Hitting Time and Place to a Sphere or Spherical Shell for Brownian Motion with Drift," *Statistics and Probability Letters*, vol. 42, no. 4, pp. 367–373, May 1999.

45. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 3rd edition, 1991.

46. EIA/TIA, "Cellular Radio-Telecommunications Intersystem Operations," *EIA/TIA, Tech. Rep. IS-41 Revision C*, 1995.

47. W. Ho, A. Yuen and W. S. Wong, "A Dynamic Location Area Assignment Algorithm for Mobile Cellular Systems," in *Proc. IEEE International Conference on Communications*, 1998, pp. 1385–1389.

48. Z. Naor and H. Levy, "Minimizing the Wireless Cost of Tracking Mobile Users: An Adaptive Threshold Scheme," in *Proc. IEEE INFOCOM Conference*, 1998, pp. 720–727.

49. C. Rose, "Minimizing the Average Cost of Paging and Registration: A Timer-based Method," *ACM Wireless Networks*, vol. 2, no. 2, pp. 109–116, 1996.

50. I. F. Akyildiz, J. Ho and Y.-B. Lin, "Movement-based Location Update and Selective Paging for PCS Networks," *IEEE/ACM Transactions on Networking*, vol. 4, no. 4, pp. 629–638, 1996.

51. S. Tekinay and A. Catovic, "Optimal Geolocation Updating for Location Aware Service Provisioning in Wireless Networks," in *Proc. IEEE Vehicular Technology Conference Spring*, 2001, pp. 2600–2604.

52. T. S. Rappaport, *Wireless Communications: Principles & Practice*, Prentice Hall, 2001.

53. P. Gupta, P. R. Kumar, "The Capacity of Wireless Networks," *IEEE Transactions on Information Theory*, vol. 46, no. 2, pp. 388–404, Mar. 2000.

54. M. Grossglauser and David N. C. Tse, "Mobility Increases the Capacity of Ad Hoc Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 10, no. 4, pp. 477–486, Aug. 2002.

55. V. Rodoplu and T.H. Meng, "Minimum Energy Mobile Wireless Networks," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 8, pp. 1333–1344, Aug. 1999.