ABSTRACT

AIA, MAHESH G. Directed Waypoint Model: A Hybrid Approach to Realistic Mobility Modeling in Mobile Ad-Hoc Networks. (Under the direction of Assistant Professor Mihail L. Sichitiu).

Researchers are highly dependent on simulations to compare and evaluate the performance of Mobile Ad-hoc networking protocols. Mobility modeling plays a very important role in simulation-based studies of MANET protocols, since the mobility of nodes is the limiting factor in the performance of such protocols. Currently, simulations of MANET protocols rely on simplistic, purely stochastic models such as the random waypoint. These models do not reflect the motion of real world MANET nodes, which may be deployed in diverse scenarios. At the present time, the only available alternative to these models is fine-grained detailed simulations based on extensive surveys of actual node behavior, or extensive sets of real world traces. This work proposes an alternative hybrid mobility model that extracts scenario-specific information from real traces and generates realistic node movements for particular scenarios. The distinguishing features of our model are its ability to reproduce spatial-temporal node movement properties from sample traces and its general adaptability to traces from any real-life scenario. We develop a framework for our model and apply it to reference traces from several realistic scenarios. The hybrid model is then evaluated for accuracy by comparing the performance of MANET routing protocols using random waypoint with realistic parameters and our model against the original reference traces used by our model. The results show the relevance of accounting for spatial and temporal aspects of node movements in mobility models.

Directed Waypoint Model: A Hybrid Approach to Realistic Mobility Modeling in Mobile Ad-Hoc Networks

by

Mahesh G. Aia

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial satisfaction of the
requirements for the Degree of
Master of Science

Department of Computer Science

Raleigh

2005

Approved By:

Dr. Yannis Viniotis

Dr. Rudra Dutta

Dr. Mihail L. Sichitiu
Chair of Advisory Committee

To my parents for their everlasting and unconditional love . . .

Biography

Mahesh Aia was born on February 11, 1981 in Daman, India. After completing his early education in Goa, he attended Manipal Institute of Technology and earned a Bachelor of Engineering (B.E) degree in Computer Science and Engineering. Upon graduation, in June 2002, he took up a position as project associate at the Indian Institute of Science, Bangalore where he worked on various projects related to cost-effective computing technologies. In Fall 2003, he joined North Carolina State University for graduate studies in Computer Networking. He has been a member of the Wireless and Local Area Networks research group ever since. During Summer and Fall of 2004, Mahesh worked fulltime as an Intern at TapRoot Systems Inc., at Research Triangle Park, NC in their Internet telephony group.

Acknowledgements

My foremost thanks to my advisor, Dr. Mihail Sichitiu for his constant guidance and support over the last two years. I am indebted to him for his innumerable ideas about the research work and his help, both technical and otherwise. Without his help and guidance, this work would never have come to fruition. The intense, and at times humorous, brain storming sessions with him are my most enduring memories of the last two years at NCSU.

I feel very thankful to Dr. Yannis Viniotis and Dr. Rudra Dutta for agreeing to be on my thesis committee.

I would like to thank all members of the WALAN group, particularly Harshvardhan Joshi, for several lively discussions related to my research, and Jangeun Jun for his help in everything from LaTeX templates to presentation ideas. I thank my friend Avesh, for always being there whenever I needed to talk to someone. My gratitude is also due to my former roommate, Venkat, for his useful criticisms and amazing company.

I am also very grateful to my sister, Dr. Pratibha, and my brother, Ajit, for helping me in ways too numerous to recount. Without their support, love and friendship, I would not be what I am today.

Finally, I reserve my highest gratitude for my parents and all my alma maters for the quality education they imparted to me. I am forever indebted to my parents, to whom I dedicate this thesis.

Contents

List of Figures vii

List of Tables viii

1 Introduction 1
- 1.1 Motivation . . . 1
- 1.2 Contributions . . . 4
- 1.3 Thesis Organization . . . 5

2 Related Work 6

3 The Directed Waypoint Model 11
- 3.1 Requirements and Limiting Factors for Realistic Mobility Modeling . . . 11
- 3.2 Mobility Characteristics Measured by Directed Waypoint Model . . . 15
  - 3.2.1 Number of Nodes . . . 15
  - 3.2.2 Node Spatial Distribution . . . 16
  - 3.2.3 Pause Durations and Locations . . . 17
  - 3.2.4 Group Mobility . . . 17
  - 3.2.5 Speed of Motion . . . 18
- 3.3 An Integrated Framework for the Directed Waypoint Model . . . 19
  - 3.3.1 Architectural View . . . 19
  - 3.3.2 Pass 1 : Trace Analysis . . . 20
  - 3.3.3 Pass 2 : Trace Regeneration . . . 22

4 Performance Evaluation 25
- 4.1 Evaluation Setup . . . 25
  - 4.1.1 General Evaluation Methodology . . . 25
  - 4.1.2 Base Scenario : Pedestrian Mobility on University Campus . . . 26
- 4.2 Effect of Simulation Parameters . . . 29
  - 4.2.1 Transmission Range . . . 29
  - 4.2.2 Simulation Time . . . 31
  - 4.2.3 Number of Nodes . . . 32

4.2.4 Traffic Load 33
4.2.5 Routing Metric 33
4.2.6 Reactive Routing Protocol 35
4.3 Performance in Different Scenarios 36
4.3.1 Scenario 1: Movement of Airplanes across the United States 37
4.3.2 Scenario 2: Real Traces of Buses in Seattle 38
4.3.3 Scenario 3: Motion of Bees around a Beehive 39
4.4 DWP Intensive Evaluation 41
4.4.1 Number of Replays 41
4.4.2 Weight Assigned to Spatial Distribution 43
4.5 Concluding Remarks 44

5 Summary and Future Work 46
5.1 Future Work 46

A Trace Analysis Algorithms 48

B Trace Regeneration Algorithms 55

Bibliography 57

List of Figures

3.1 Design space of a hybrid mobility model 14
3.2 Directed Waypoint Model architecture 19
3.3 The method of replays 23

4.1 Transition diagram for a node on the university campus 28
4.2 AODV routing performance as a function of transmission range 29
4.3 AODV routing performance as a function of simulation time 31
4.4 AODV routing performance as a function of the number of nodes 32
4.5 AODV routing performance as a function of traffic load 33
4.6 AODV average end-to-end delay as a function of the number of nodes 34
4.7 AODV routing control overhead for various number of nodes 35
4.8 Bellman Ford routing performance as a function of the number of nodes . . . 36
4.9 AODV routing performance for airplanes in the US 38
4.10 AODV routing performance for buses in Seattle 39
4.11 AODV routing performance in beehive scenario 40
4.12 AODV routing performance as a function of the number of DWP replays . . 42
4.13 AODV performance for varying spatial distribution weights 43

List of Tables

4.1 Default simulation parameters for DWP Evaluation 28
4.2 AODV routing performance for default parameters 29

Chapter 1

Introduction

1.1 Motivation

Owing to the advances in wireless communication technologies and their cost effectiveness in recent years, advanced wireless networking technology has seen increasingly widespread use in diverse applications. Mobile Ad hoc Networks (MANETs) [1] are a special class of wireless networks in which a group of autonomous mobile platforms –herein simply referred to as nodes– cooperatively forward packets over wireless links, without relying on any external infrastructure. In traditional applications of wireless networking, such as cellular technology and wireless local area networks, only the last hop between the core network backbone and the end user node is a wireless link. MANETs on the other hand, must forward packets from one node to another in multihop fashion because the nodes themselves are responsible for routing packets.

MANETs have applications in areas such as military operations, disaster relief and impromptu collaborative computing where it may not be feasible to have access to network infrastructure. Sensor networks, vehicular ad hoc networks and mesh networks are a few of the more recent applications of MANETs. The developing technologies of ubiquitous computing and communications may provide new applications for MANETs. There are likely several potential applications of MANETs that are not presently realized or envisioned.

Nodes of a MANETs may move arbitrarily, causing frequent disruptions and changes in routes. Highly dynamic, random and sometimes rapidly changing topologies are, therefore, an inherent feature of these networks. The high level of nomadicity brings with it several new challenges that are not present in the design and evaluation of either traditional, hardwired packet networks or infrastructure-based wireless networks. For instance, routing protocols rely on topological information to operate correctly. Consequently, their performance is severely affected by the degree and nature of node mobility in the MANET [2, 3, 4]. Likewise, node mobility has an effect on TCP performance [2], design of topology control protocols for energy conservation [5] and possibly even on the design of encoding channels at the physical layer by virtue of frequent changes in the signal-to-noise ratio. Therefore, MANETs warrant research into the design of mobility-aware networking protocols at every level of the protocol stack.

Performance evaluation by simulation is an important tool that networking researchers use to measure the effectiveness of proposed protocols. Simulations allow for the creation of hypothetical scenarios, replicable experiments and rapid exploration of the design space of a particular solution. Since MANETs have not been widely deployed at this time, detailed information regarding the real-time performance of specific protocols is difficult to obtain. Hence, simulations are particularly important in evaluating MANET protocols. Given the pervasive influence of mobility on MANET protocols, the problem of generating node movements that closely mirror real world situations assumes critical importance in any simulation-based evaluation of MANET protocols.

The problem of generating realistic node movements does not have a trivial solution. The exact behavior of the nodes may vary from one scenario to another, and within each scenario from time to time. Moreover, there are geographic constraints such as obstacles and pathways that influence node motion. Node behavior is often highly scenario-specific and hard to generalize. The vision of mobile ad hoc networking is to design protocols that scale to very large autonomous networks involving thousands of nodes over areas of several square miles. Future MANETs may be deployed in scenarios as diverse as rescue personnel in a disaster relief operation, sensors attached to wild animals, cars on a freeway or even aircraft in flight. In this context, it is unacceptable to incorporate any assumptions specific to the deployment scenario within a mobility model. In short, an ideal mobility model must generate realistic node movements while being generic enough to be applicable to any given real-world scenario. The model must in essence be adaptable to any given scenario. Research on the effect of mobility on routing protocols [3] validates this assertion.

Mobility models being used at the present time fall well short of meeting these requirements. Currently used models fit into two broad paradigms. The stochastic models are based on some rather simple assumptions and are general in nature. Their ease of implementation makes them highly popular. However, they have no provision to adapt to specific scenarios. For instance, the performance of a MANET protocol in a battlefield may be markedly different than on a university campus for identical simulation parameters. However, the statistical behavior of nodes in a given stochastic model follows a fixed, predictable pattern [6, 7]. The parameters associated with stochastic models are insufficient to reflect the wide variation in node behavior observed across scenarios. Such models based on the concept of random walk have been popular in the cellular networking research community to model the worst case (completely random) motion of nodes around the base station. However, in the context of MANETs, such models have resulted in inconclusive and often conflicting results [6, 8]. The relative performance of MANET protocols vary according to the specific stochastic mobility model used [6]. This could be attributed to the fact that mobility modeling in MANETs, unlike in infrastructure based wireless networks, amounts to modeling not only the motion of the nodes, but also the routing functionality incorporated within them. Randomized mobility is too generic and not representative of the way MANET nodes move in real world scenarios. Consequently, routing protocol evaluation based on such unrealistic mobility models may lead to wrong conclusions being drawn about their performance, and render their practical applicability highly suspect.

Detailed simulation models on the other hand model node movements from a particular scenario to a very high level of accuracy. However, a detailed simulation is tied to a specific instance of a scenario. Such simulations require very fine-grained real-world information about node behavior and extensive computational power to store and manipulate this information. For instance, a realistic simulation of the mobility of cars and pedestrians in city would require a detailed study of pedestrian and vehicular traffic behavior over several days. We would need details such as the specific lane rules in force, schedules and working hours of establishments in the area, the proportion of the population which commutes to work and so forth. Thus if we were to evaluate the general behavior of a protocol by this method, we would need access to all the associated low-level information and high computational resources to run the simulation multiple times with different settings. Due to these logistical complexities, detailed simulations are rarely used to evaluate MANET protocols. In this thesis, we implement a few such simulations, although at a much less detailed level, for comparison purposes.

Given the shortcomings of both these methods of mobility modeling, a hybrid model encompassing the strengths of both paradigms suggests itself as an intuitive and promising alternative. Recent research on the role of mobility in MANET routing protocol performance [3] indicates that the spatial and temporal dependencies between node movements is a decisive factor. Hence a model with the ability to reproduce the spatial and temporal relationship in a given scenario would be capable of generating realistic movements without the need for highly detailed information entailed by detailed simulation based studies.

1.2 Contributions

This thesis presents Directed Waypoint Model (DWP), a hybrid mobility model that automatically adapts to a specific scenario given a sample set of node movement traces from the scenario. We first propose a set of metrics that capture the defining properties of the reference mobility scenario. Next, we formulate and implement a framework to automate the extraction of relevant statistical information from the input traces and then apply the statistics so obtained to generate synthetic traces in accordance with user-specified parameters. Unlike detailed simulations, DWP does not require fine-grained scenario specific information and nor is it tied to a specific instance of the general mobility scenario. The only input to the model is a set of recorded node movement traces from the reference scenario. The source of these traces could be real-time measurements such as the location traces available for buses in Seattle [9] or the sample output of a detailed simulation conducted by specialized organizations [10]. DWP differs from purely stochastic models such as Random Waypoint in two crucial aspects: first, the node destinations, speeds and pause times are selected from distributions derived from real data as opposed to uniform distributions between user-specified limits. Second, DWP captures the spatial and temporal dependencies between node movements from the given traces and makes a best-effort attempt to reproduce the same in the generated traces. In essence, DWP represents a middle ground between the completely general but unrealistic stochastic models on one hand and the realistic but highly specific detailed simulations on the other. We prove the reliability of our model by comparing the performance of the AODV routing protocol [11] using DWP against the results obtained using the input traces as well as the Random Waypoint Model. The main advantage of DWP is that, given real traces from any scenario, it provides a realistic and reproducible method to evaluate the performance of the protocol in real life.

1.3 Thesis Organization

The thesis is organized as follows : Chapter 2 describes the related work. Chapter 3 discusses the requirements of realistic mobility modeling and describes the design of the Directed Waypoint model along the details of trace analysis and extraction algorithms. Chapter 4 compares and contrasts the performance evaluation of MANET routing protocols using Directed Waypoint Model as opposed to purely stochastic models such as Random Waypoint Model. Finally, Chapter 5 summarizes our findings and discusses related future work.

Chapter 2

Related Work

Since mobility models are an essential component of any network simulator to evaluate mobile networking protocols, a number of them can be found in the literature. Existing work fits broadly into two distinct paradigms: purely stochastic models and detailed simulation frameworks. In general, stochastic models are based on rather simple assumptions regarding the motion of nodes. Their advantage is that they are analytically tractable and allow the calculation of mathematical expressions to characterize protocol performance. Detailed simulation based studies on the other hand describe the motion of the nodes in a detailed and realistic manner, but in general do not allow the derivation of analytical expressions.

A comprehensive survey paper [6] classifies stochastic mobility models as entity and group mobility models [6]. Of the entity mobility models, the random walk, random waypoint, random direction and boundless simulation models all simulate completely randomized and memory-less mobility patterns. Random walk model basically produces Brownian motion. The random direction mobility model forces mobile nodes to travel to the edge of the simulation area before changing speed and direction. The boundless simulation model uses a closed coverage area that may be visualized as a torus. The Gauss-Markov mobility model uses a tuning parameter to vary the degree of randomness. Similarly, the probabilistic mobility model [12] and city section mobility model [13] use probability distribution functions to control randomness. Among group mobility models, the exponential random correlated (ECR) model [14] reproduces all possible node movements by adjusting the parameters of a motion function. However, it is not easy to correlate the values of these parameters and distributions with available real-world information.

By far the most commonly used, purely stochastic, model in ad-hoc networking research is the Random Waypoint Model [15]. In this model, a mobile node selects an arbitrary destination within the simulation area and a speed that is uniformly distributed between *[minspeed,maxspeed]*. The node then proceeds to the destination at the chosen speed. Upon reaching the destination, the node pauses for a time period drawn from a uniform distribution within pre-specified limits. The process repeats until the end of the simulation. The pause duration along with the speed of node motion controls the nomadicity of the mobile nodes.

A number of authors have pointed out statistically undesirable properties associated with the Random Waypoint Model and attempted to rectify the same by introducing improved derivatives. Hong and Gerla [14] have noted that in models like the random waypoint model, the choice of node speed and direction are not correlated to previous values, resulting in sudden speed changes and sharp turns. To remedy these shortcomings, Bettstetter [16] describes *Smooth Random Mobility Model* in which the change in speed and direction of node are controlled by stochastic processes that induce more realistic node behavior. Bettstetter and Wagner [7] note the non-uniform spatial node distribution produced by random waypoint model and suggest a modified model known as *Random Borderpoint Model* that achieves a more uniform node distribution. It is a modified version of random waypoint model in which the selected destinations can only be located on the borders of the simulation area. Camp, Boleng et al [6] note that the random initialization of nodes in random waypoint model may produce high variability in results in the first few seconds of the simulation until steady state is achieved. To fix this initialization problem, Navidi and Camp [17] derive steady state distributions of the random waypoint model for node location, speed and pause time. When the initial locations and speeds of the nodes are chosen from these stationary distributions, convergence to steady state is immediate.

Recently, many researchers have recognised the drawbacks of purely stochastic models for determining MANET performance in real world environments and have proposed specialized models to accurately capture the dynamics of specific real world scenarios. In general, the proposed models introduce geographic constraints that act as barriers to node motion and signal propagation. The layout and specifications of these constraints rely on assumptions about the particular scenario being modeled.

Cavilla, Baron et al [18] provide empirical evidence to show that the generalized models like random waypoint and even specialized simulation models for indoor MANETs fail to realistically model protocol performance in an obstacle-rich environment. The graph-based mobility model proposed by Tian, Hahner et al [19] uses a graph to model the movement constraints faced by nodes in city. The obstacle mobility model [20] attempts to create more realistic node movements through the incorporation of obstacles. Obstacles are represented as arbitrarily complex polygons that are specified by user input. The objects serve as barriers to both node movement and wireless transmission between the nodes. Pathways are computed as the edges of a Voronoi graph with the obstacles as the vertices. The distinguishing feature of these models is that nodes are constrained to follow designated pathways. Although these models generate more realistic node movements than purely stochastic models, they address only one of the problems of such models namely the representation of geographic constraints. Besides, the representation of obstacles requires detailed information about the specific scenario being modelled in the form of geographic maps.

Johansson et al [4] have conducted a scenario-specific evaluation of ad hoc routing protocols. Various scenarios corresponding to different levels of node mobility and network conditions are pre-defined, and then the performance of the protocol in each of these scenarios is evaluated. The drawback of this approach is that a separate scenario needs to be defined for each combination of network conditions and node mobility levels against which protocol performance is to be measured. However, given the diverse conditions in which MANETs are likely to be deployed in future, pigeon-holing each real-life scenario to one of a small set of pre-defined hypothetical ones does not represent a general solution.

The TRANSIMS [10] framework follows the detailed simulation paradigm. It generates accurate and realistic node movements by using highly detailed and specific scenario information based on an extensive survey of node movement habits. Such low-level information is rarely available for most scenarios and hence, it is not suitable for researchers seeking feedback on the performance of their protocols under diverse conditions.

Hollick et al [21] present a mobility model for wireless metropolitan area networks which is a hybrid of an empirical mobility model and a synthetic traffic model. It is a macroscopic model developed in the context of transportation and land use modeling. In this model, *trips*, which define the movement of users from an origin to a destination are mapped onto *zones*, which define areas with a certain attraction level determined by socioeconomic characteristics. The generated trips are distributed among the destination zones according to the attraction of these zones. User preference for locations is characterized by classifying users into disjoint categories such as consumer, worker, trainee etc. While this approach is appropriate for metropolitan area networks, where nodes are primarily human beings inhabiting a city, the granularity of scenario-specific detail inherent in this model makes it unsuitable for mobile ad hoc networks with diverse deployment scenarios.

Musolesi et al. propose a novel group mobility model based on realistic models of human socialization [22]. The authors use mathematical models of human socialization to group hosts together in a way that is based on social relationships among the individuals. Tan et al [8] present an individually simulated behavioral model that simulates the movement behavior of individual nodes and their relationships to surrounding nodes in order to generate the overall mobility pattern. However, both these approaches are only applicable when the mobile nodes are being carried by walking human beings, so that the movement of nodes is influenced by human perception and socialization behavior.

The IMPORTANT framework [3] seeks to objectively evaluate the effects of mobility on the performance of MANET routing protocols by benchmarking them against a set of various protocol-independent mobility metrics. The results confirm the intuition that node mobility is a very important factor influencing performance. Further, it proves that conclusions about the relative performances of specific protocols are inconsistent, and vary according to the mobility model used. This paper also identifies the aspects of MANET routing protocols which affect their performance in a mobile scenario. However, the authors do not suggest a general mobility model that generates realistic movements for all possible scenarios.

Cano, Manzoni et al. [23] highlight the significant impact of group mobility on the performance of MANET routing protocols. In particular, the mix of inter- and intra-group traffic is shown to strongly influence routing performance. The authors introduce four different models that combine random waypoint with the concept of group and compare the performance of the DSR routing protocol using these models as well as random waypoint. The high variability of the results reaffirms the need for a mobility model that can adapt itself to the prevailing spatial-temporal dependencies of a given scenario.

The work most closely related to ours can be found in Hsu et al [24]. This work proposes a hybrid approach similar to ours, but is considerably more limited in scope. The authors rely on manual observation and surveys to gather and analyze data from a campus scenario. This work is preliminary in nature and specific to a campus scenario. DWP on the other hand, incorporates a much larger set of metrics to measure mobility scenarios and completely automates the analysis and extraction of the relevant statistics from the reference scenario. Moreover, DWP is designed to be inherently general, without making any assumptions about the scenarios where it might be applied. We argue that a truly hybrid mobility model must be completely automated and general purpose in nature to be usable in diverse, yet to be envisioned scenarios.

Chapter 3

The Directed Waypoint Model

3.1 Requirements and Limiting Factors for Realistic Mobility Modeling

Most evaluations of MANET protocols so far have used mobility models based on *random walk* of mobile objects, henceforth referred to as purely stochastic models. By far the most commonly used among these models is the random waypoint model (RWP)[15]. Since the RWP model is highly representative of the class of stochastic mobility models and also is the most commonly used, we use it as a benchmark for comparisons in this thesis. Unless otherwise qualified, statements about the random waypoint model apply equally to other purely stochastic models.

As mentioned earlier, RWP provides two configuration parameters, namely the range of the speed distribution and the range of the pause distribution. During the simulation, the specific values for node speeds and pause times are selected from a uniform distribution between the respective limits. The advantage of RWP is its simplicity and generality. By varying the range of node speeds and pause times, we can simulate different rates of network topology change. However, RWP makes a number of simplifying assumptions that are not valid in real-life MANET scenarios. For instance, in the real world the node population varies dynamically with time. Likewise, node speeds and pause times do not necessarily follow a uniform distribution, or for that matter any other well-known mathematical distribution. Node density is usually not uniform either, with some hotspots attracting a larger proportion of the nodes. Most importantly, in real MANETs the rate of topology change is not a simple function of speed and pause times. MANETs work in an inherently cooperative environment. Since the cooperation of peer nodes in forwarding each other's traffic is the defining feature of a MANET, the spatial and temporal dependency between the location of mobile nodes has a direct correlation with the frequency and duration of route disruption [3]. The tendency of mobile nodes to move in groups, known as group mobility, is a manifestation of such spatial-temporal dependency, and is known to affect MANET routing protocols [23]. To further complicate matters, some studies show that the rate of topology change itself may not completely determine protocol performance [8]. While the exact relationship between mobility and protocol performance is still an open problem at this time, it is evident that the configuration parameters provided by models like RWP are too simplistic and insufficient to reflect the complex dynamics of real world MANETs.

There is clearly a need for a mobility model that factors in such properties that have an impact on protocol performance. However the difficulty with accounting for these properties is that they are typically scenario dependent and therefore it is not possible to evaluate them *a priori* in the absence of specific real world data. Clearly, incorporating all low level information about various scenarios cannot be the basis of a generic mobility model. A generic model must be able to deduce relevant scenario-sensitive characteristics and adapt itself given only basic high-level data. The most basic form of real world data is a trace of recorded space and time coordinates for mobile nodes in a particular scenario, which are generally hard to obtain, given the sparse deployment of real ad hoc networks at the present time. Even when such real data is available, it is usually specific to the particular circumstances prevailing at the time of the recording and therefore not representative of the scenario in general. For instance, location traces of 100 MANET nodes moving on a particular university campus during an arbitrary time period cannot be directly used to gauge the expected performance of a MANET protocol in a campus scenario because many of the trace characteristics are dependent on the time of recording. Nor can it be used to evaluate the performance of the protocol in a hypothetical scenario with different parameter values such as the number of nodes or the speed of movement. At the present time, the only feasible method to obtain an accurate generalized evaluation is to run detailed simulations that take into account all the details associated with node movement in the scenario. In case of the university scenario, such details might include the lane rules followed by cars and pedestrians, class schedules and detailed maps of the campus among others. Such detailed information about specific scenarios is rarely available readily, therefore it requires extensive surveys and studies. Since detailed simulations are based on very large data sets, generating node traces involves an uncommon amount of computational power. Given these limitations, detailed simulations are not easily accessible as a method for MANET protocol evaluation.

Thus, we observe a rather sharp divide between the highly general but simplistic purely stochastic models on one hand and highly specific but expensive detailed simulations on the other with no middle ground. Given that very few MANETs have been deployed in real-time, the availability of real traces itself is traditionally viewed as a hard requirement to satisfy. Hence any evaluation that relies on real traces is considered infeasible for practical purposes. However, with recent advances in ubiquitous sensing technology [25] and the proliferation of mobile platforms equipped with geo-location facilities such as GPS [26], we argue that availability of real traces of potential MANET nodes is no longer an unrealistic assumption. Furthermore, sample results of detailed simulations conducted in specialized scenarios are often available in the public domain. However, given sample traces, we are still left with the problem of deducing general node movement properties to evaluate expected protocol performance for the scenario. In this context, we focus on developing a hybrid stochastic mobility model that extracts relevant information from given node traces and uses it as a hint to generate more accurate and scenario-sensitive node movements than RWP.

It is convenient to think of the properties associated with a given trace of MANET node movements as falling into two distinct categories. Some properties such as the pattern of node arrivals and departures, the tendency of nodes to move in groups, speed of motion and duration of pause intervals are *inherent* to a particular scenario. For instance on a university campus, the pattern of student arrivals or the location and dimensions of hotspots such as libraries and classrooms tend to remain largely invariant from day to day. Other properties such as the number of nodes or the number of hotpots are *temporal*. The number of students on campus may vary from day to day and during a given day from time to time. Likewise, on a busy day, we observe a larger number of *hotspots* since more classroooms are occupied.

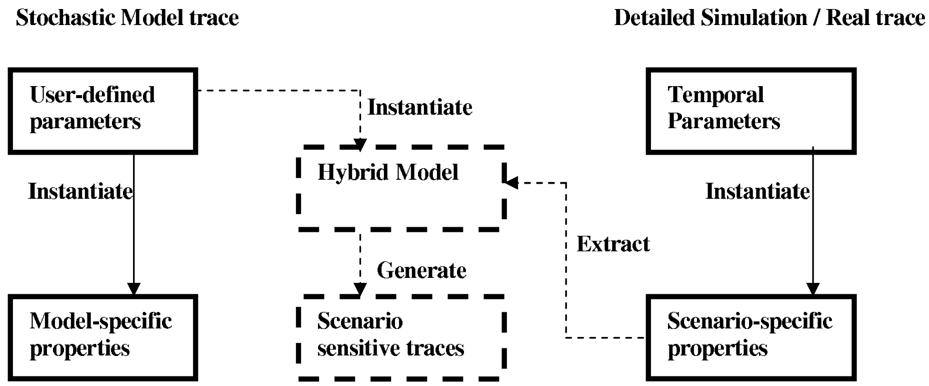

Figure 3.1: Design space of a hybrid mobility model

With this abstraction in mind, we view a trace as consisting of a core of immutable inherent properties *instantiated* by a mutable set of values corresponding to the temporal parameters. The problem of predicting general protocol performance in the reference scenario therefore reduces to accurately identifying, characterizing and measuring the inherent characteristics from the traces. Having done so, we can evaluate the general performance of the protocol by generating traces for different combinations of temporal characteristics. We are thus able to accurately evaluate MANET protocol performance for specific scenarios without having to know the low-level details of that scenario. Figure 3.1 illustrates the design space occupied by such a hybrid approach.

The accuracy of the hybrid approach is limited by the extent and granularity of the traces as well as the computational power at our disposal. When large sets of sample traces are available, the accuracy will approach that of a detailed simulation provided we have sufficient computational resources to process all the traces. With reduced computational power, we may choose to sample the traces at periodic intervals rather than process all the recorded coordinates with a resultant trade-off in accuracy. With a very limited set of input traces, the accuracy will approach that of purely stochastic models.

It must be noted that our view of realistic mobility modeling is limited to the accurate characterization of key spatial-temporal dependencies between node movements and other properties associated directly with the input traces. In our view, a set of generated traces may be deemed realistic if the performance of a protocol with these traces closely models the performance on the reference input traces themselves. That is, the generated traces can be realistic even if a neutral observer does not perceive any difference in the trajectories between directed waypoint model and other regular purely stochastic models. In particular, we do not attempt to model the effect of geographic constraints and obstacles on protocol performance because such information cannot be obtained from a simple trace of node movements. Other researchers have attempted to model these factors [20, 18] and such work may be combined with our view of realistic mobility modeling to increase the overall accuracy of generated traces.

3.2 Mobility Characteristics Measured by Directed Waypoint Model

The first step in formulating a hybrid framework for realistic mobility modeling is to identify the key inherent properties in the given input traces that must be reproduced in the output synthetic traces. In this section we identify these properties, discuss their relevance to realistic mobility modeling and suggest metrics to quantize and measure them. To do so, we systematically consider each of the sources of non-realism in purely stochastic models like RWP.

3.2.1 Number of Nodes

In most real world MANETs, the node population in an area of interest varies with time. Node density is highly correlated with protocol performance. A high number of nodes typically translates into a higher average number of neighbors per node, which directly influences the route availability. However, models such as RWP use a fixed number of nodes for the duration of the simulation. The nodes are initially randomly distributed around the simulation area. This initial random distribution of mobile nodes may not be representative of the manner in which nodes distribute themselves while moving. Depending on the simulation time, this effect can produce high variability in the results [6].

To eliminate these anomalies, we need to deduce the arrival and departure pattern of nodes from the given traces, irrespective of the actual number of nodes observed. Note that the arrival and departure pattern itself is an *inherent* property whereas the number of nodes is *temporal*. We define the following three quantities : *bulk arrival size* is the number of nodes that arrive together. *Inter-arrival time* is the time between between successive bulk arrivals. *Node Lifetime* is the amount of time a node spends between its arrival into the simulation area and its exit. The first two quantities characterize the node arrival pattern, while the third dictates the time when node exits the area.

All of the above quantities may be measured as stationary probability distributions of the respective occurrence counts observed in the real traces.

3.2.2 Node Spatial Distribution

In purely stochastic models, the long-term node distribution of nodes over the simulation area is invariant for a given model. In RWP for example, nodes tend to concentrate towards the center of the simulation area [16]. The aim of such models is to achieve uniform node distribution. In real world MANETs on the other hand, node distribution varies from scenario to scenario. This is because nodes in real world scenarios usually travel to well-defined destinations that carry important semantic meaning, such as a classroom, library and so on. The long-term distribution of nodes over the simulation area affects the connectivity of the topology graph and hence the routing performance. In this context, assuming a uniform node distribution or an invariant distribution peculiar to the mobility model in use can lead to unrealistic results.

For our purpose, we define a hotspot as a region where many location measurement samples are clustered together in the recorded trace. A hotspot could either be a region through which many nodes pass through, or a region where nodes pause for long durations. In order to accurately characterize the spatial node distribution present in a scenario, we must be able to locate the hotspots from the given traces. Next, we must measure the relative popularity of each hotspot in relation to others. Hotspots can be characterized by stationary distributions of their areas and the number of measurements recorded within the area enclosed by them. The second characteristic is an indicator of a hotspot's relative popularity.

Note that while the exact number of clusters and the measurements enclosed within them are *temporal* properties specific to the recorded traces, their relative distributions by area and popularity may be classified as *inherent* properties.

3.2.3 Pause Durations and Locations

In RWP, nodes pause for a duration selected from a uniform distribution between pre-specified limits. Since real world MANETs rarely have uniformly distributed pause times, the average pause time even if known, can be a misleading statistic. The distribution of pause times has a bearing on the stability of the network topology and thus affects protocol performance. Besides the distribution of pause times, the locations where the nodes pause can also be an important attribute, especially in scenarios with long pause durations.

The solution is to measure and use the stationary distribution of pause times as observed in the recorded traces. In order to control the locations where nodes pause, we require that nodes restrict their choice of destinations to the hotspot clusters identified for the node spatial distribution. In addition to a measure of the relative popularity of the hotspot, we can also associate with each cluster a *pause probability, p*, the ratio of consecutive identical recordings over the total number of recorded measurements within the cluster. When a node reaches its destination, it decides to pause with probability p associated with the hotspot or not pause with probability $1-p$. If it decides to pause, it selects a pause duration from the stationary distribution of pause times.

Since individual nodes choose their pause durations independent of all other nodes, the stationary pause time distributions measured from the input traces are representative of generic node behavior in the scenario and can be considered an *inherent* property. Similarly, while the exact number of paused coordinates recorded within a cluster is *temporal*, the pause probability itself may be considered *inherent.*

3.2.4 Group Mobility

In RWP, as in other entity-based pure stochastic models, nodes move completely independent of each other. It therefore does not capture the tendency of nodes to move in groups, a phenomenon often observed in real life scenarios. Group mobility can be an important characteristic affecting routing protocol performance depending on the mix of inter and intra group traffic [23]. While models such as random waypoint group model are designed to simulate group mobility, they are prone to using unrealistic values for group sizes and number of groups. Nodes in real world MANETs seldom move in well-defined groups with a constant number of group members within a fixed distance of each other.

We view group mobility as a spatial cum temporal dependency. Abstractly, any two nodes are said to belong to the same group if they are consistently within some threshold distance D of each other for at least an interval of time T. We herein refer time threshold as the *persistence time*. The number of nodes that belong to a group is known as *group cardinality*. Groups of nodes in real world MANETs are not necessarily well-defined. That is to say, they do not necessarily have well-defined and constant values of D and T. Besides, several groups may overlap each other. It is therefore difficult to pinpoint the prevalent group cardinalities from the recorded traces. In order to simplify the characterization of group mobility, we do not attempt to measure the exact group cardinalities themselves. Instead for a user-provided estimate of the *close enough* distance for the scenario, we measure the stationary distribution of observed persistence times.

Again, while the exact numbers corresponding to the persistence times depend on the number of node pairs in the recorded traces and are therefore *temporal*, the relative distributions of various persistence times are *inherent*.

3.2.5 Speed of Motion

Like in the case of pause times, RWP selects node speeds from a uniform distribution. When speeds are selected from a uniform distribution with low minimum speed, then at any given time a large proportion of nodes will be moving very slowly, creating a near stable network topology that results in unrealistically good performance [27].

As with pause times, the remedy is to measure the stationary distribution of speeds observed in the real traces. Node speed in itself does not ordinarily depend on number of nodes and can therefore be considered *inherent*.

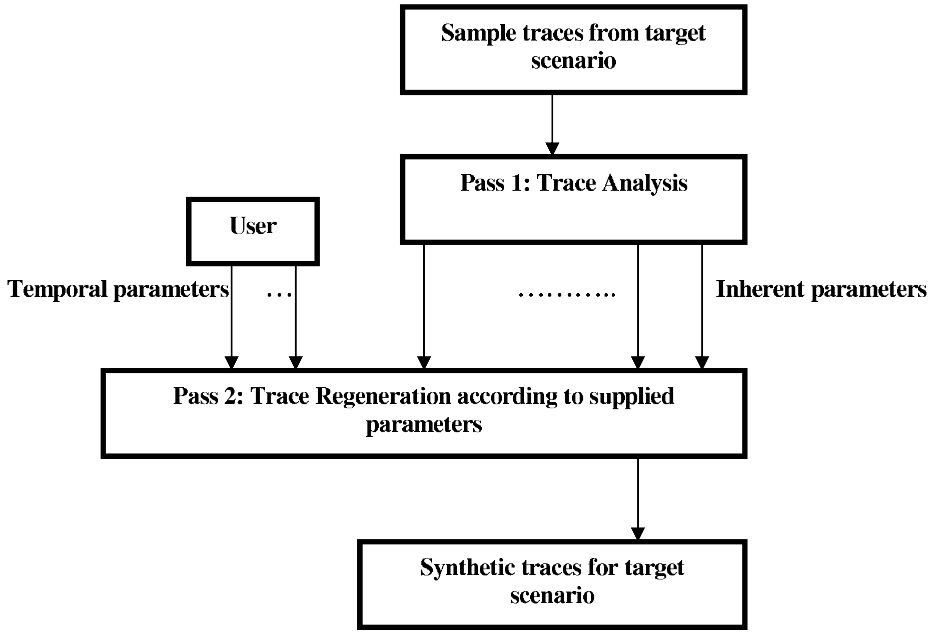

Figure 3.2: Directed Waypoint Model architecture

3.3 An Integrated Framework for the Directed Waypoint Model

3.3.1 Architectural View

As seen in figure 3.2, Directed Waypoint Model consists of two forms of input : A set of traces from the real life scenario and user-supplied information. The operation of the model proceeds in two distinct phases. In the *trace analysis* phase, we perform *post-facto* analysis of the given traces to extract stationary distributions of inherent properties as identified in section 3.2. Next, in the *trace regeneration* phase, we normalize the distributions with user-defined temporal parameters to define the target distributions for the current instantiation of the model. Once the target distributions are identified, the process of actually generating the traces is similar to Random Waypoint Model, except that destination selection and node motion are governed by the target stationary distributions : nodes arrive on the scene in accordance with the target bulk size and inter-arrival time distributions. On arrival, nodes also select an expiry time. Each node selects a destination from the list of designated hotspots, and then proceeds to that destination at a pre-determined speed. Upon reaching the destination, a node decides to pause with the associated pause probability for that hotspot and selects a corresponding pause duration. When the pause duration comes to an end, the cycle repeats. On reaching its expiry time, the node departs the simulation area.

The inherent properties are derived from the given traces, whereas the temporal parameter values are supplied by the user. However, the user has the ability to override any inherent information if he so desires. This allows for the creation of hypothetical scenarios. While overriding, the user may modify or altogether replace the stationary distribution associated with an inherent property. Conversely, if the user does not provide temporal parameters, information from the input traces will be used by default. This provides a means for testing the accuracy of the generated traces.

3.3.2 Pass 1 : Trace Analysis

In this section we describe the algorithms and techniques to measure the stationary distributions corresponding to the properties identified in section 3.2. The algorithm details in pseudo-code form may be found in Appendix A.

The input traces are of the form *(xcoords[1 ... n][1 ... t], ycoords[1 ... n][1 ... t])* where n is the number of nodes whose coordinates have been recorded and t is the duration of the recording. We assume that node positions are available at regular intervals of time. A zero value indicates that the node is not active for the particular time instant. If the actual traces available are not in such a form, we may use interpolation functions to generate approximate recordings at the required time instants.

Algorithm *surfconc* pre-processes the input coordinates to facilitate subsequent identification of hotspots by algorithm *find-hotspots*. In particular, algorithm *surfconc* generates a list of locations found in the traces, along with the occurrence count for each location. It also identifies the subset of location coordinates at which measurements have been recorded in successive time intervals. The last two quantities are known as the *coordscount* and *pausecount* distributions respectively.

Algorithm *find-hotspots* identifies hotspots in the recorded traces and computes relevant statistics to characterize them. The input to the algorithm is the set of locations computed by algorithm *surfconc*, along with distance and time parameters, d and t respectively. It uses the idea of time-based clustering presented by Kang et al [28]. The original algorithm is meant to extract significant places visited by a single user from a trace of node movements. Since we are concerned with the more general problem of determining the prevalent node spatial distribution in the scenario, we generalize the algorithm to work on the collective set of traces from all nodes. Any two recorded locations are assumed to belong to the same cluster, as long as their separation in space and time is within the distance and time thresholds respectively. Finally, all overlapping clusters are merged together. Information about each cluster is quantized by a two-dimensional histogram representing the area of the cluster and the sum of *coordscount* for all locations within the cluster. The ratio of the total *pausecount* to the totalcoordscount for all included locations represents the *pause probability* for the cluster.

The value of the distance threshold parameter d can significantly affect the size and number of clusters. If the size of d is large compared to the dimensions of the reference area where the traces are recorded, the size of clusters may become large and absorb the intermediate locations where not many nodes have been recorded. A small value for d on the other hand entails several small fragmented clusters instead of a single cluster of larger dimensions. Likewise, the parameter t controls the tolerance of the algorithm to temporal separation between traces. A large value of t may result in semantically unrelated locations being clustered together. A small value on the other hand can lead to locations being assigned to disjoint clusters despite being semantically related. Intuitively, the value of the time threshold is correlated with the speed at which nodes move. A high rate of node movement implies a smaller tolerance to temporal separation and conversely, slow moving nodes necessitate a larger value of t.

Algorithm *perstime* computes the persistence time distribution as defined in section 3.2. We divide time into a number of evenly spaced *epochs*. The persistence time is computed over each successive *epoch* in a cumulative fashion. That is, the persistence time distribution for the final epoch e subsumes the persistence time information for all preceding *epochs*. The necessity of epochs arises from the temporal nature of persistence times, and will become clear when we discuss the *Trace Regeneration Pass*. Although this algorithm uses a distance threshold d, similar to algorithm *find-hotspots*, it is important to note the nuance in the semantic meaning of the two parameters. The persistence time between nodes is measured as the duration for which the nodes consistently lie within a distance d of each other. In the hotspot identification algorithm, d denotes the tolerable spatial separation for any two recorded locations to belong to the same cluster.

The remaining algorithms measure the stationary distributions for speed, pause-time, node lifetime, inter-arrival times and bulk arrival sizes. These algorithms are based on straightforward parsing of the recorded traces to measure the desired quantities.

3.3.3 Pass 2 : Trace Regeneration

Having measured the inherent characteristics in the reference traces, we now address the problem of generating realistic traces that possess similar characteristics.

We begin by randomly designating specific clusters within the simulation area as hotspots. The area, target *coordscount* and *pause probability* for each cluster is selected from the two-dimensional histogram computed in Pass 1. The number of clusters is a temporal parameter supplied by the user. Next, we schedule the arrival times of the nodes in accordance with the arrival size and inter-arrival time distributions. Once the clusters and arrival schedules are in place, the nodes begin moving as outlined in section 3.3.1.

One aspect of producing realistic traces is to select the values for stochastic variables such as pause durations, node speeds and arrival times from the stationary distributions obtained. However, modeling spatial-temporal dependencies is not as straightforward. Since spatial-temporal properties are global in nature, their accurate reproduction requires concerted action on the part of all nodes. DWP, on the other hand, is essentially an entity based mobility model where the motion of each node is independent of the movement of other nodes. In order to account for such inter-nodal properties while still preserving the entity-based semantics of DWP, we introduce the *method of replays*, which essentially is a form of closed-loop feedback control. We allow the nodes to move and pause on their own accord. At discrete intervals of time, known as *checkpoints*, we compare the spatial and temporal properties of the traces generated so far with the target distribution at the same time, namely the *coordscount* and *perstime* distributions. To increase the chances of finding a good match, we can generate multiple *replays* between checkpoints. The *replay* which results in the best match with the target distribution so far is committed as the final set of traces for the interval. Figure 3.3 illustrates the idea. The concept of cumulative *epochs* introduced earlier fits naturally into this scheme. An *epoch* corresponds to the time interval between *checkpoints*. The reason behind computing persistence times on a per-epoch basis is to provide a fair basis for comparing generated traces at the intermediate checkpoints.

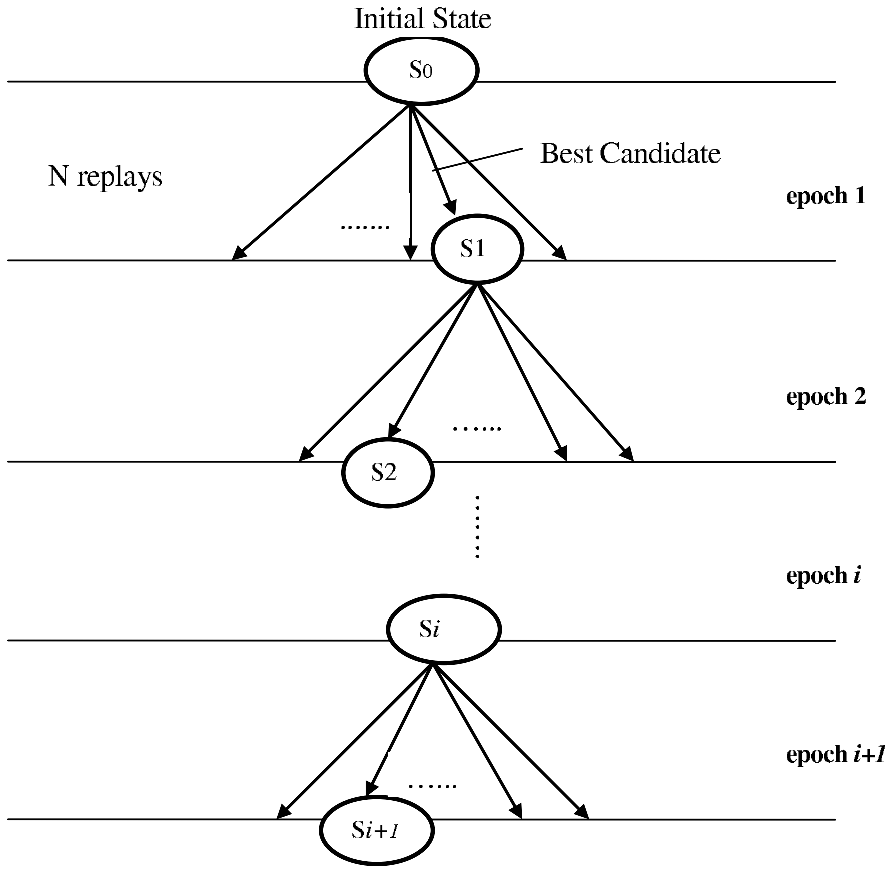

Figure 3.3: The method of replays

The *coordscount* distribution is assumed to be uniformly distributed over all epochs. The length of each *epoch* corresponds to the time period between successive checkpoints, and therefore affects the granularity of feedback control exercised by the model. Ideally, the epoch length should be sufficiently large to observe significant differences in the spatial temporal properties of the candidate replays. Very long epochs on the other hand result in replays with widely varying properties. A reasonable heuristic is to choose the epoch length long enough to cover the entire journey of a node between successive pauses over the worst case distance of the simulation area.

The target distributions for the spatial-temporal properties are affected by the relative number of nodes in the measured traces and the number of nodes for the current instantiation, which is a user-defined temporal parameter. If n is the user-specified number of nodes and N is the number of nodes in the reference traces, the *coordscount* distribution is normalized by the factor n/N while the *grouptimes* distribution is normalized by $n(n-1)/N(N-1)$.

Algorithm *comparereplays* compares the candidate replays for an epoch and returns the one which best matches the current target spatial and temporal property distributions. The pseudo-code details are provided in Appendix B. It uses algorithm *perstime* to compute the persistence time distribution corresponding to each replay. Besides, the *coordscount* for each replay is also computed. The sum of the absolute differences between the target persistence time and achieved persistence time distribution, *perstime* is weighted equally with the sum of absolute differences between the achieved and targeted spatial distribution, *coordscount*, for all clusters to arrive at a *score* for each replay. Note that the target *coordscount* and persistence time distributions are on different scales and hence must be normalized before being added together. The normalization factor we use is simply the ratio of their maximum possible target values, which is the sum of the *coordscount* for all clusters and the sum of the values for all persistence time histogram intervals respectively .

It must be noted that DWP attempts to reproduce spatial-temporal properties on a best-effort basis. The accuracy of the spatial-temporal characteristics of the generated traces with respect to the target distribution depends on the number of *replays*. For a given set of reference traces, as the number of *replays* increases, we would expect the accuracy of the generated traces to improve.

Chapter 4

Performance Evaluation

4.1 Evaluation Setup

4.1.1 General Evaluation Methodology

In this section we evaluate the reliability and general applicability of DWP by using it to predict the performance of AODV [11], a commonly used MANET routing protocol. The goal is to determine how closely the performance trends predicted by DWP match those in the detailed simulation traces; we perform a similar comparison with the commonly used RWP model. We compare the performance of AODV with three different sets of node traces with identical CBR network traffic patterns between 50 randomly selected node pairs. We choose CBR instead of TCP, in order to separate the effects of TCP's congestion control mechanisms from the effects of node mobility. The benchmark set of traces are directly generated by detailed simulations of real-life scenarios. The two candidate traces are RWP with parameter values matching those from the detailed simulation, and traces generated by the DWP framework with the detailed simulation traces as input. We generate the first set of traces by conducting coarse-grained detailed simulations to mimic the motion of nodes in various real-life scenarios. These traces also serve as input to the DWP framework implemented in Matlab, which generates one of the sets of candidate traces. DWP is used in *comparison* mode with the temporal parameters being the same as those observed in the real traces. The user does not override any information, so that the resulting traces are directly comparable. Traces for RWP model are generated with the *setdest* program [29] in NS2, with the mean pause time and maximum speed parameters matching the values observed in the detailed simulation traces. The three sets of traces are then fed to the NS2 network simulator and performance of AODV is evaluated against each of them. All other simulation parameters, including CBR traffic patterns are identical for all three evaluations. Traffic patterns are between randomly selected source-destination pairs beginning at arbitrary times within the initial 3000 seconds of the simulation. The experiment is repeated 30 times with different connection patterns. The simulation results are plotted along with their 95% confidence intervals. We then compare the results from DWP with the results obtained when using RWP with equivalent parameters and the actual reference traces themselves. Data packet delivery ratio (PDR) is the primary metric used in our comparisons. We also measure end-to-end delay and routing control overhead.

We first define a base reference scenario with default simulation parameters. We then independently measure the effects of varying these simulation parameters and DWP's general applicability by varying the reference scenario itself. Finally, we conduct an intensive evaluation of DWP's fundamental features.

For our evaluation purposes, we choose to implement our own detailed simulations to generate the benchmark traces for the most part, rather than use real traces. We do so because real traces are not available at this time for a variety of scenarios. Since reliable information about the prevailing spatial dynamics in the reference traces is available to us, analysis of the performance trends predicted by DWP vis-a-vis the reference traces and RWP model also becomes more tractable. It must be noted that DWP itself is agnostic to the origin of the traces. In order to validate the applicability of DWP to real world traces, we also conduct an evaluation of DWP with real node traces of buses moving in Seattle.

4.1.2 Base Scenario : Pedestrian Mobility on University Campus

In our base reference scenario, we consider the mobility of pedestrians during a typical working day on NCSU historical campus. To generate the reference traces, we implement a coarse-grained detailed simulation which mimics the motion of students on a 1000 x 1000 m section of campus over a 14 hour period between 7AM to 9PM. All locations on campus are abstracted as either classrooms or leisure spots and the user must specify the coordinates of all these locations. We specify these values to closely match the layout of buildings on NCSU historical campus. The user also specifies the entry and exit points, the frequency of node arrivals and the range of arrival sizes of nodes. We assign the typical frequency and capacity of university transportation buses to these variables. The maximum speed of node motion is set to 3 m/s with a probability distribution of 0.8 for the interval 0-1 m/s, 0.1 for 1-2 m/s and 0.1 for 2-3 m/s. The detailed simulation framework reads in the university schedule for classes, known as TRACS schedule [30], along the information specified by the user. After determining the number of nodes required to satisfy the enrollment figures in the schedule, arrival times and class schedules are generated for each node. Once the node arrives, it proceeds to a class or a leisure spot depending on its schedule. The node pauses in the classroom until the end of class and then proceeds to a leisure spot if there is sufficient time or to another classroom otherwise. Upon completion of its schedule, the node departs the simulation area. The state transition diagram of a node in the campus scenario is depicted in figure 4.1.

Upon completion of the simulation, the generated traces are fed as reference to the Directed Waypoint Model. The distance thresholds for algorithm 2 and algorithm 3 are set to 10m. In case of RWP model, the mean pause time parameter is set to the expected value of the pause time distribution calculated by the trace analysis phase. Similarly, the maximum speed parameter is set to the maximum value in the speed distribution. The simulation area dimensions match those in the detailed simulation. The MAC protocol used in the evaluation is 802.11 with default settings.

Pedestrian mobility on campus represents a scenario in which nodes move at relatively low speeds, with a wide range of pause durations between successive trips. Nodes also have a high tendency to move together and cluster in classrooms and other university buildings. Another feature of this scenario is that nodes arrive and leave the campus periodically. The practice of switching off the network-enabled handheld devices is also a manifestation of the same phenomenon.

The default network simulation parameters are shown in table 4.1. Table 4.2 indicates the routing performance of AODV for these settings with the three candidate traces from the campus scenario. The DWP framework runs 50 replays per epoch.

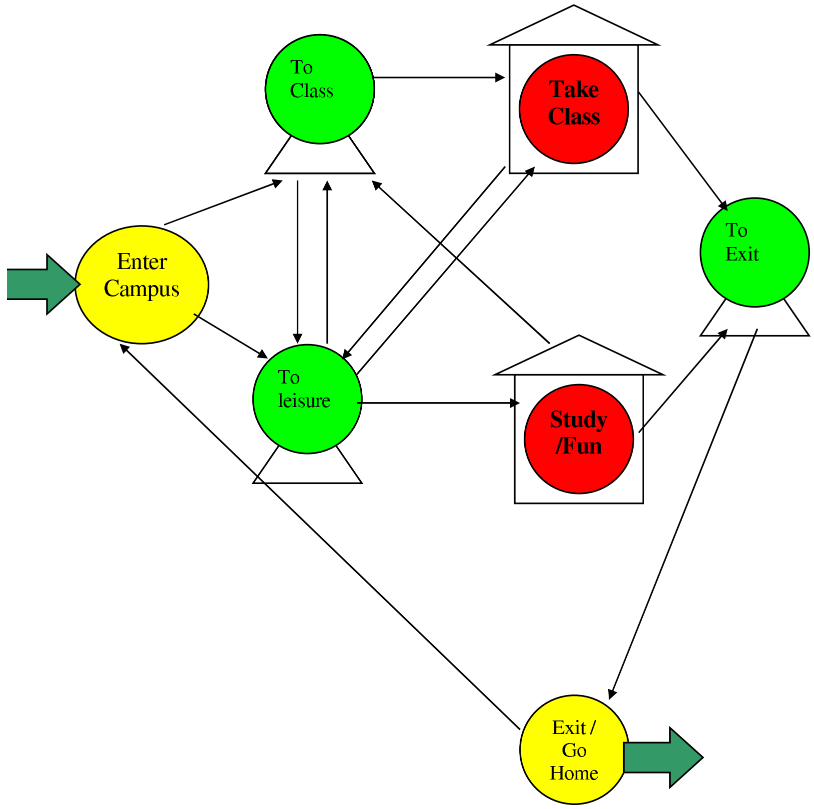

Figure 4.1: Transition diagram for a node on the university campus

Table 4.1: Default simulation parameters for DWP Evaluation

Parameter	***value***
Transmission Range	200m
Simulation Time	25000s
Number of Nodes	75
Routing Protocol	AODV
Metric	PDR
Traffic load	50 connections
Traffic Type	CBR
Traffic Rate	2 packets/s
Packet Size	512 Bytes

Table 4.2: AODV routing performance for default parameters

	RWP	***DWP***	***Simulation Traces***
PDR%	98.39	70.41	82.31

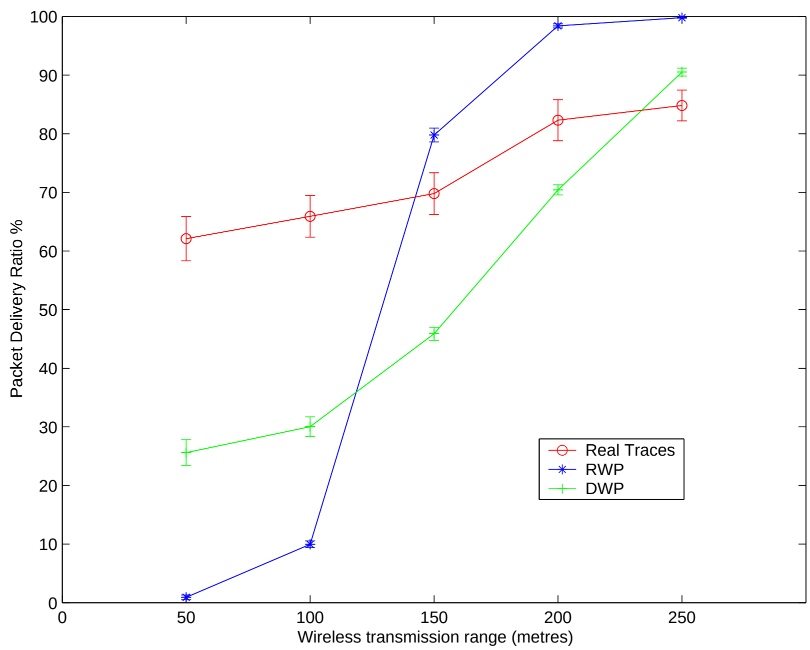

Figure 4.2: AODV routing performance as a function of transmission range

4.2 Effect of Simulation Parameters

In this section, we evaluate the effect of simulation parameters on the accuracy of performance results indicated by DWP with respect to the reference traces and RWP. To do so, we vary one parameter at a time while all the remaining parameters retain their default values.

4.2.1 Transmission Range

In the first experiment, we evaluate the routing performance of AODV with increasing radio transmission range of nodes, varying from 50m to 250m insteps of 50. The results are shown in figure 4.2.

As nodes are stationary for most of the time in this scenario, we expect the number of one-hop neighbors to increase with transmission range, resulting in markedly higher ratio of packets being successfully delivered. From 4.2, we see that while the data PDR does improve with transmission range in all cases, the increase is very gradual in the real traces. In case of the RWP model on the other hand, the PDR is very low for smaller transmission ranges. However, small increases in transmission range result in steep improvements, until around 200m range where the PDR is close to 100 percent. In the directed waypoint model, the PDR although lower than that observed in the real traces, is better than RWP for smaller transmission ranges. We observe that the slope of the DWP graph is intermediate between the extremes represented by the real traces and the RWP model. At higher transmission ranges, the performance projected by DWP closely matches the observed values from the real traces.

The steep increase in performance while using RWP traces is explained by the fact that the long-term spatial node distribution of RWP is close to uniform. Consequently, an increase in transmission range results in an almost quadratic increase in the number of one-hop neighbors per node, given that nodes are stationary most of the time. This is not necessarily the case in a real-life campus scenario, since students spend most of the time seated together in classrooms and other buildings. The increase in transmission range does not have a significant impact once the range extends beyond the dimensions of the buildings in which nodes happen to be located. This also explains the relatively good routing performance observed in the real traces even for small transmission ranges. Going back to the detailed simulation used to generate the reference traces, we observed that the dimensions of most of the buildings specified were within 100m. The reason the slope of the DWP graph is steeper than the real traces, but significantly less so than RWP is that node behavior in DWP resembles the behavior of nodes in the real environment in some, but not all respects. Like the real environment, nodes in DWP can select only designated hotspot clusters as destinations. Likewise, pause durations are selected from the distribution obtained from the reference data. However, the difference is that the selected pause durations are not correlated with the node's current position, since DWP is agnostic to the semantic meaning of specific locations. Whereas in the real scenario, the nodes enter the classrooms at the beginning and remain there for the entire time period of the class, in DWP, nodes enter and leave in a disorganized fashion. Therefore, while DWP does simulate hotspots, it results in a less stable network topology than the real environment.

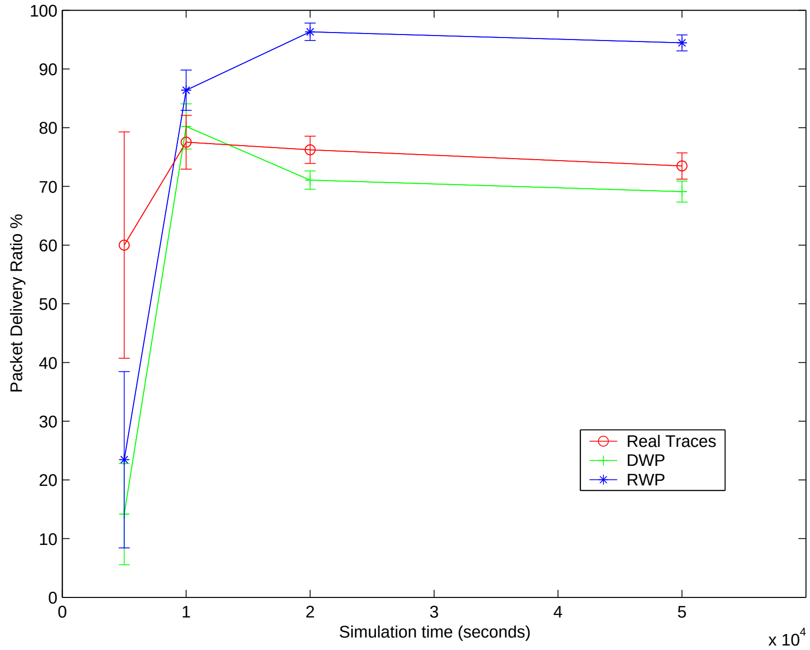

Figure 4.3: AODV routing performance as a function of simulation time

4.2.2 Simulation Time

We now vary the simulation time from 5000 seconds to 50000 seconds and observe the variation in the indicated performance. The results are shown in figure 4.3.

We observe high variability in the results for the first few experiments. There are two reasons for this phenomenon. Firstly, the CBR traffic sources begin sending traffic at arbitrary times within the first 3000 seconds. Hence, the number of connections as well as the duration for which they have been active vary widely across connection patterns. Secondly, in case of the reference traces and DWP, all nodes have not yet arrived in the simulation area within the first few thousand seconds. As a result, the number of recipients of the traffic sources varies across connection patterns.

After the initial transition period, all three sets of traces show almost constant performance for the remainder of the simulation. The routing performance indicated by the DWP traces closely matches the reference traces after the initial 10000 seconds. RWP, on the other hand, indicates a much higher PDR than the reference traces throughout the simulation. Once RWP reaches its steady-state, almost uniform spatial node distribution,

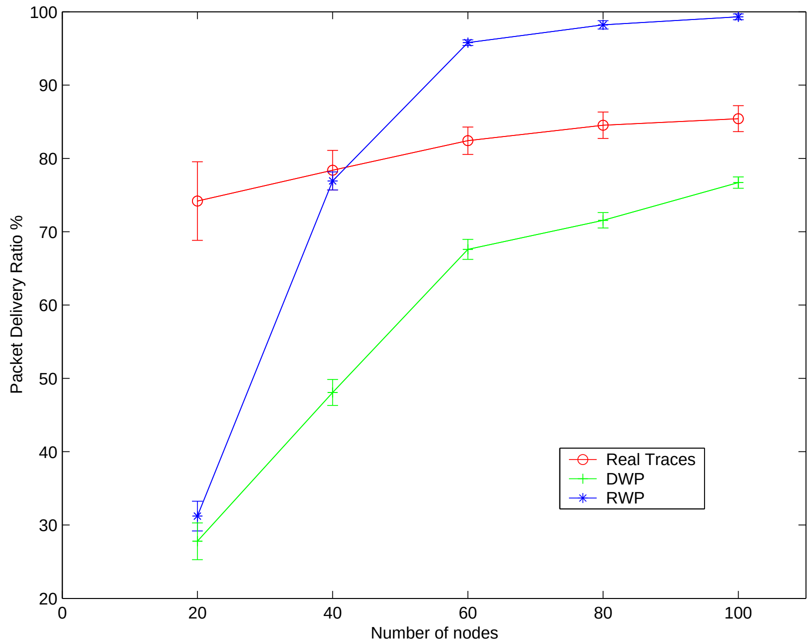

Figure 4.4: AODV routing performance as a function of the number of nodes

the relative immobility of the nodes together with the relatively high transmission range result in unrealistically good routing performance.

4.2.3 Number of Nodes

In this experiment, we vary the number of nodes participating in the simulation from 20 to 100 in steps of 20. The results are shown in 4.4.

Like in the case of transmission range, we expect the routing performance to increase with number of nodes. Again, the performance improvement is gradual for real traces and steep for RWP. The performance projected by DWP closely parallels the real traces at high number of nodes. Since students are seated together in classrooms for most of the time, adding more nodes to an already dense local topology results in only marginal improvement in the ratio of packets delivered. Again, due to the almost uniform node distribution observed in RWP, a larger number of nodes directly translates into better route availability throughout the network thereby indicating a rapid improvement in performance. When the number of nodes is greater than 60, the spatial distribution of DWP begins to

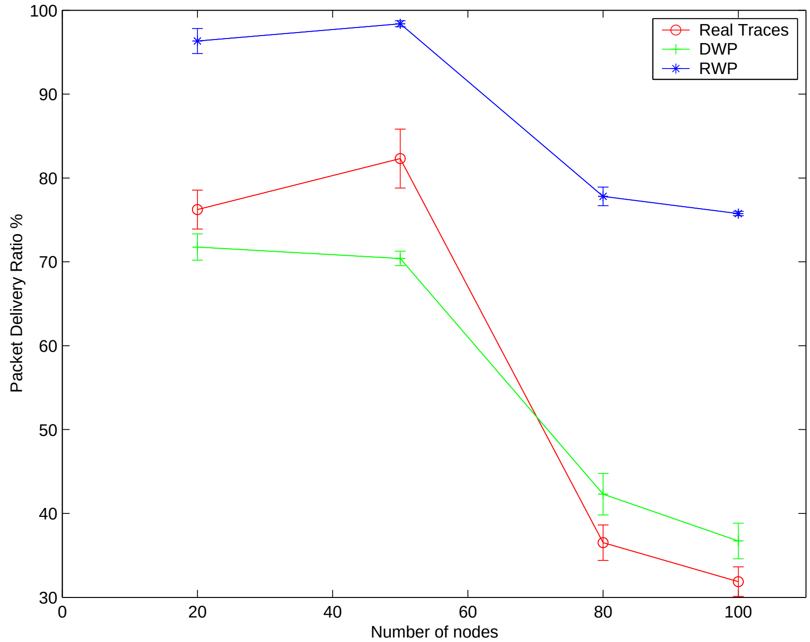

Figure 4.5: AODV routing performance as a function of traffic load

approach that of the real environment, resulting in similar improvements in performance.

4.2.4 Traffic Load

We evaluate the effects of traffic load on the accuracy of DWP by varying the number of CBR connections from 20 to 100, while keeping the number of nodes constant at 75. The results are shown in figure 4.5.

We note the sharp decline in the ratio of successfully delivered packets in the real traces when traffic load increases. The DWP performance trend closely mirrors the real traces. RWP, however, scales unrealistically well with traffic load. The decline in performance is much less pronounced in this case. The spatial reuse accruing from the almost uniform node distribution is the most plausible explanation for this trend.

4.2.5 Routing Metric

We now measure the routing performance in terms of average end-to-end delay and routing control overhead, rather than the Packet Delivery Ratio. As in the preceding

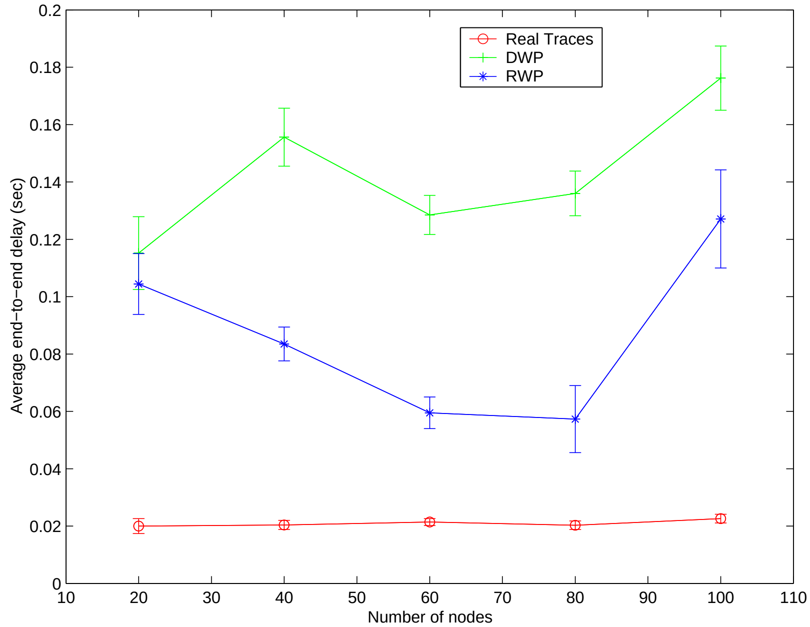

Figure 4.6: AODV average end-to-end delay as a function of the number of nodes

experiment, we vary the number of nodes participating in the simulation from 20 to 100 with all other parameters being the same as in the default case. The results are shown in 4.6 and 4.7.

We note that the end-to-end delay remains almost constant in the reference traces. This is due to the fact that nodes are stationary for most of the time, confined within buildings. Moreover, the nodes enter and exit the buildings in lockstep fashion. Hence, a high proportion of the packets which get delivered are likely within the building in which the node happens to be located, independent of the number of nodes participating in the simulation. In case of both, DWP and RWP, we do not observe any regular pattern in the end-to-end delays, although the absolute values are significantly higher than observed in the reference traces. In case of RWP, the uniform node distribution is a contributing factor to the high end-to-end delay. Since the nodes are uniformly distributed, packets get delivered across several hops. Although DWP simulates hotspots, the entry and exit of nodes to these hotspots is not as coordinated as in the reference traces. Hence, packets get delivered both within and beyond the designated hotspots resulting in inaccurate end-to-end delay statistics.

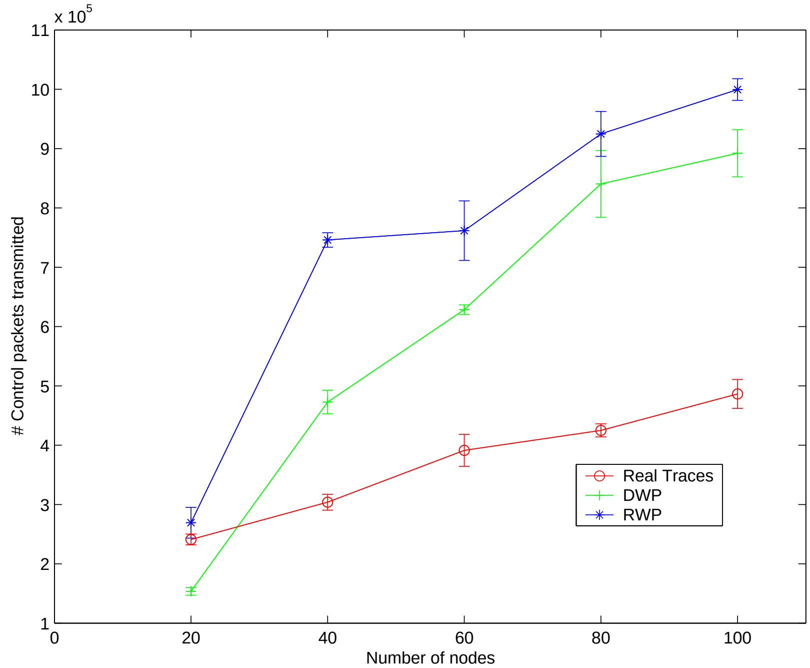

Figure 4.7: AODV routing control overhead for various number of nodes

In case of routing control overhead, we note the relatively gradual increase in the number of control packets sent out in the real traces. Both,RWP and DWP, to a smaller extent, indicate significantly higher overhead. The highly stable and localized topology resulting from the real traces is most likely the reason behind this phenomenon.

4.2.6 Reactive Routing Protocol

We next evaluate the performance results for a different routing protocol. We select the Bellman Ford routing algorithm, which differs fundamentally from AODV in that it is a proactive routing protocol. The results are shown in figure 4.8

As in the case of AODV, we note that the routing performance remains almost constant with the number of nodes. Likewise, DWP performance values are much less than actually observed for low number of nodes, but closely mirror the real traces for 60 and more nodes. RWP, on the other hand shows significantly different performance trends for Bellman Ford as compared to AODV. RWP starts at a much higher PDR than earlier for small number of nodes. When additional nodes participate in the simulation,

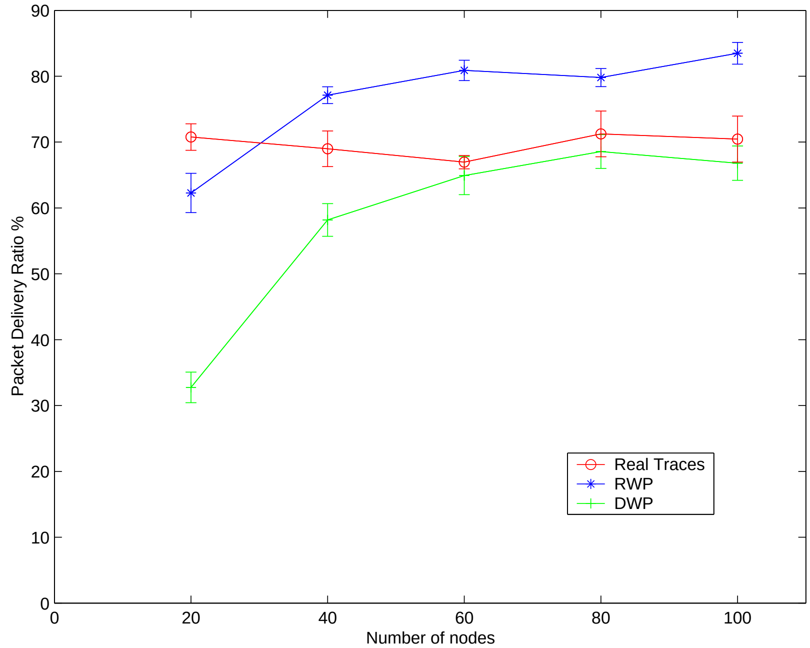

Figure 4.8: Bellman Ford routing performance as a function of the number of nodes

RWP performance does not rise as sharply or to the same levels as earlier. These results indicate that spatial and temporal properties of node movements may have similar effects on fundamentally different MANET routing protocols. This trend is not accurately captured by RWP.

4.3 Performance in Different Scenarios

In this section, we evaluate the accuracy of DWP performance results for different realistic scenarios. As far as possible, we retain the same simulation parameters as those shown in Table 4.1. However, some parameters such as the radio frequency and wireless transmission range need to be changed to accommodate situations peculiar to some scenarios. Such changes to the default parameters have been indicated wherever appropriate.

4.3.1 Scenario 1: Movement of Airplanes across the United States

Our first scenario represents a high level of node mobility over a wide scale. We generate traces for a hypothetical ad hoc network consisting of airplanes in flight and on the ground communicating with each other through their radios.

We implement a detailed simulation in a reference area of 4000 x 4000 km, to approximate the length and breadth of the United States. 13 airports are designated, and up to 100 planes fly directly between the airports at a speed of 672 km/h. The turnaround time between successive flights is 1-2 hours. The DWP distance parameter is set to 100 km. For the routing protocol evaluation, we use a TDMA MAC scheme with wireless communications in the Very High Frequency (VHF) range and radio transmission range of 500 km. We generate traces for a 12 hour period. The traces are two-dimensional with the Z dimension considered insignificant as compared to the other two. The protocol simulation time is 25000 seconds corresponding to the initial time period in the detailed simulation of the same length.

The distinguishing features of this scenario are the relatively sparse node density, very high speed and moderately long pause durations spread over a range of 1-2 hours. To evaluate the effect of node density and spatial distribution on AODV routing performance, we vary the number of nodes participating in the simulation and measure the PDR. The results are shown in figure 4.9.

From figure 4.9 we observe that the packet delivery ratio improves with the number of nodes for all three sets of traces. As in the case of the campus scenario, RWP indicates low PDR for small number of nodes, and subsequent increase in number of nodes result in sharp improvements in routing performance. However, the increase in RWP PDR with number of nodes is not as steep as in the case of the campus scenario. This is possibly because the high level of node mobility to some extent amortizes the uniform increase in routing capacity across the simulation area.

It is notable that the performance indicated by the DWP traces closely matches the reference traces right throughout. This is unlike the campus scenario, in which the DWP performance approaches that of the real traces as the number of nodes increases. This may be attributed to the fact that the campus scenario involves more concerted action on the part of the nodes and offers a higher degree of freedom to the nodes in terms of the pause durations as compared to the airplane scenario. The airplanes fly between airports

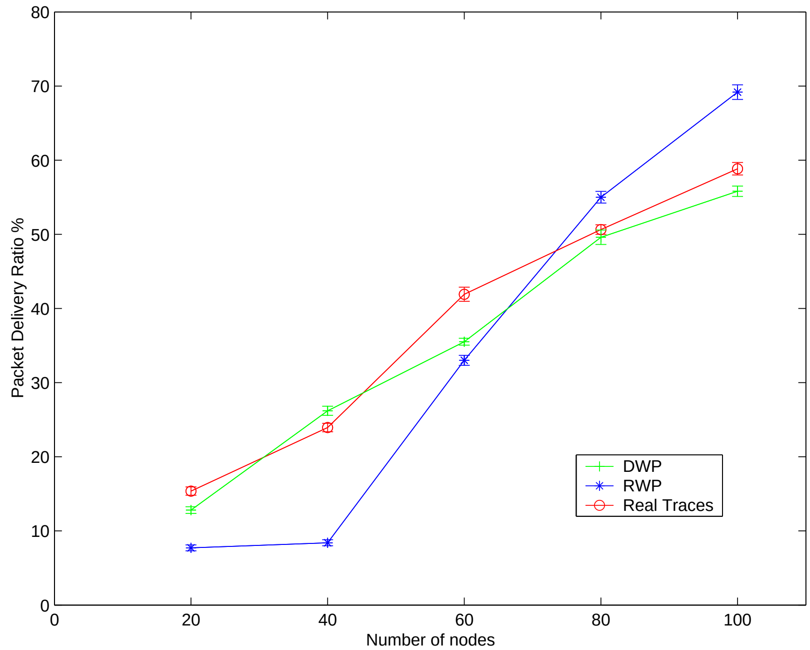

Figure 4.9: AODV routing performance for airplanes in the US

of uniform sizes and always pause between 1-2 hours. The flights are between arbitrary airports at independently determined times, unlike the campus scenario where multiple nodes move to and pause within the classrooms simultaneously, in accordance with the university schedules. Therefore, DWP is able to emulate the highly stable network topology of the campus scenario only when the number of nodes is high. In the airport scenario on the other hand, the accurate characterization of hotspots along with the pause time and speed statistics is sufficient to model performance with a high degree of accuracy.

4.3.2 Scenario 2: Real Traces of Buses in Seattle

We now focus our attention on a set of real traces of bus movements in Seattle, first used by Jetcheva et al [9] for research in metropolitan area wireless networks. The exact spatial dynamics of this scenario are opaque to us. Nevertheless, DWP trace analysis output indicates that these traces have been recorded over an area of 55 x 80 km with a maximum node speed of 27 m/s. The pause durations are typically around 3 - 4 minutes. The traces have been recorded over a period of 19 hours, with each node reporting its position every 27

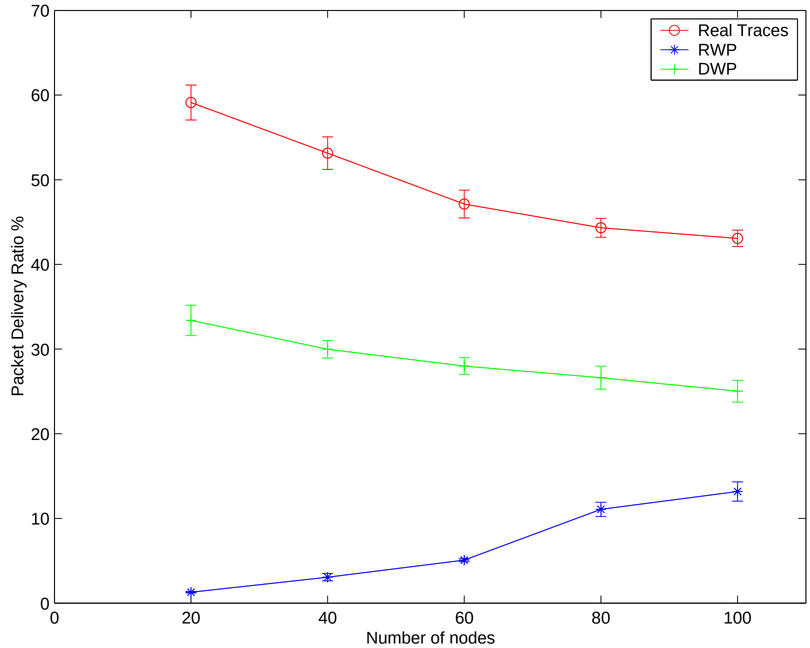

Figure 4.10: AODV routing performance for buses in Seattle

seconds. The location recordings are not synchronous; Hence we use interpolation functions to approximate the node positions at regular intervals. For the protocol simulation, we use a radio transmission range of 5 km in conjunction with a TDMA MAC scheme.

Figure 4.10 shows the results of our evaluation.

We observe that the real traces indicate a gradual decline in routing performance with number of nodes. DWP closely mirrors the trend, although there exists a significant difference in absolute values. However, DWP turns out to be far more realistic than RWP, which indicates the opposite trend of a gradual increase in PDR, along with a much larger difference in absolute values. These results indicate that DWP is equally applicable to real world traces as with detailed simulation traces.

4.3.3 Scenario 3: Motion of Bees around a Beehive

We now seek to evaluate the performance of AODV in a scenario with a moderate degree of mobility. We consider the typical behavior of bees around a beehive. A typical day in the life of a bee consists of collecting nectar from nearby flowers and then depositing

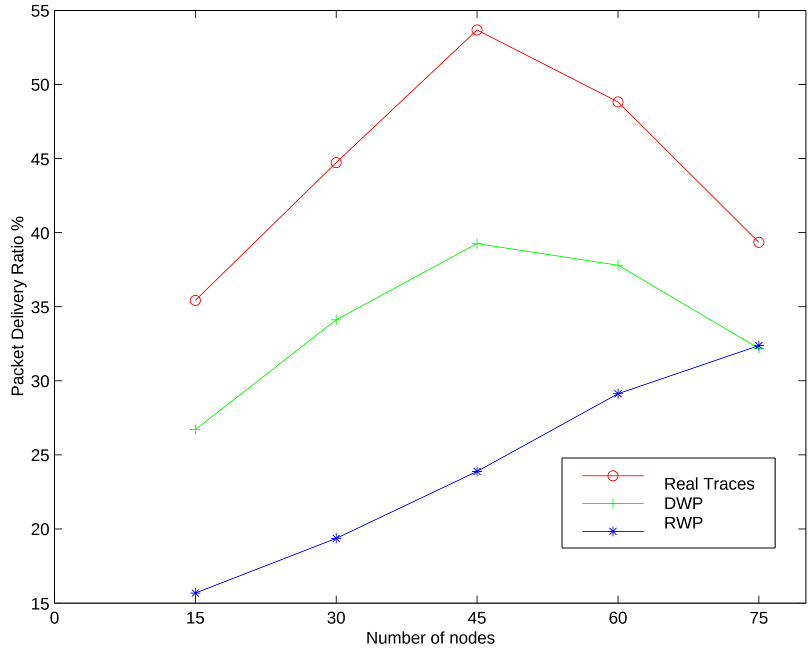

Figure 4.11: AODV routing performance in beehive scenario

it in the hive. The cycle repeats throughout the day with intervening periods of rest.

We implement a detailed simulation to mimic the behavior of 75 bees around a beehive for a 12 hour period from dawn to dusk. A field of flowers lies within a distance of 500m from the hive. Bees constantly commute between the hive and arbitrarily selected flowers in the field. The pause durations vary depending on whether the bees are located in the hive or on a flower. The typical durations are under one minute. Bee flying speeds are around 32 km/h. The simulation area is 500 x 500 m and the simulation time corresponds to the initial 25000 seconds of the detailed simulation. The radio transmission range is fixed at 75m.

The distinguishing features of this beehive scenario are the short pause durations and moderate speed of node movement. In order to evaluate the effect of node density in such a scenario, we vary the number of nodes participating in the simulation and measure the resulting performance in terms of data packet delivery ratio. Figure 4.11 shows the results.

The performance trend indicated by the reference traces in figure 4.11 shows that the routing performance initially increases with the number of nodes, and then begins to decrease. It is interesting to note that the trend predicted by DWP is very similar to the real traces, even though there is substantial disparity in the absolute delivery ratio values. RWP model on the other hand, indicates an almost linear increase in PDR with node density.

These results may be explained from the perspective of the spatial dynamics peculiar to this particular scenario. A single beehive serves as the hub of bee activity. Bees commute to and from the hive to flowers in the field. The motion of the nodes is restricted to the zone between the field and the hive. Given the relatively small transmission range, when the number of nodes increases beyond a point, the increase in routing capacity gets negated by the increased control overhead resulting from more routes being disrupted. DWP indicates a similar phenomenon. However in this case, the hotspots are fragmented and spread all over the simulation area and not restricted to the zone between the hive and the field, as DWP is agnostic to this level of detail. This explains the difference in absolute values between DWP and the reference traces. In RWP on the other hand, increase in node density is uniform across the entire simulation area. Therefore, congestion buildup is slow and fails to neutralize the increased route availability throughout the network. However, even in this case, the increase in routing performance is much slower than observed in earlier scenarios.

4.4 DWP Intensive Evaluation

Having performed an extensive evaluation of DWP by varying the network simulation parameters and testing with various realistic scenarios, we now perform an intensive evaluation of DWP's fundamental features.

4.4.1 Number of Replays

We have seen that the distinguishing feature of DWP, apart from using realistic stationary distributions, is its attempt to recapture the original spatial temporal relationships between node movements, primarily through hotspot identification and the *method of replays*. The *method of replays* was designed to emulate concerted motion of nodes within the entity based framework of DWP. Since the *method of replays* in essence amounts to a brute force search of the reference scenario's sample space, additional replays should im-

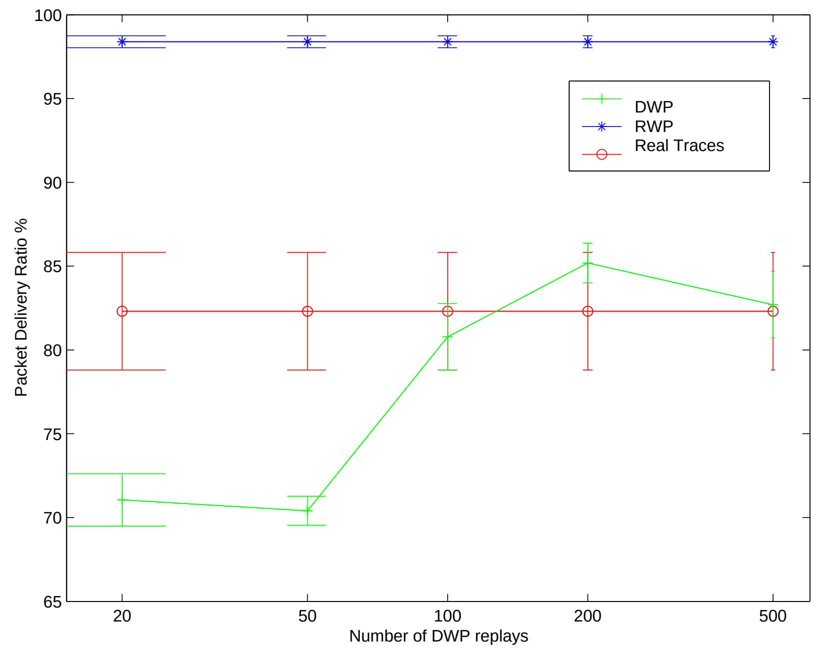

Figure 4.12: AODV routing performance as a function of the number of DWP replays

prove the accuracy of the generated traces. In order to validate this conjecture, we vary the number of replays executed by DWP for the base scenario and evaluate the traces so generated with the default network simulation parameters shown in table 4.1. We then compare the performance results with the base values in table 4.2. In order to isolate the effect of varying replays from the spatial layout of hotspots, we save the locations of the hotspots generated for the base scenario and reuse these for each run of DWP. The results are shown in figure 4.12. The PDR for RWP and the reference traces correspond to the base values in table 4.2 and are shown for comparison purposes.

We observe that as the number of replays increases, the PDR indicated by DWP approaches that of the real traces. Although the improvement in accuracy is not uniform, the indicated PDR, on the whole, tends to converge towards the actual value. In the context of the university campus scenario, we may think of the improved accuracy in the indicated PDR value as resulting from a more stable network topology. As the search space increases, the replay corresponding to the best match causes more nodes to pause in close proximity of each other for similar durations. However, the topology does not become unrealistically stable as in the case of RWP, since nodes are only allowed to pause in designated hotspot

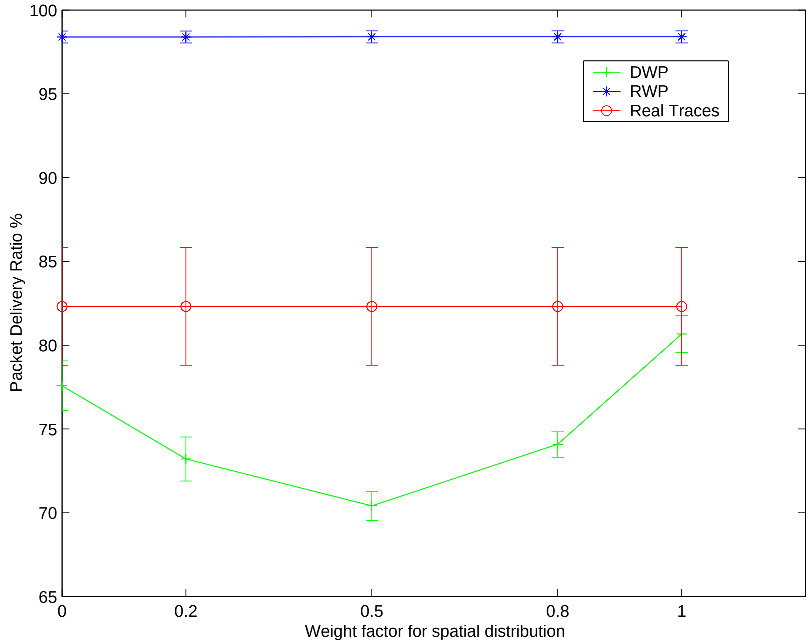

Figure 4.13: AODV performance for varying spatial distribution weights

areas. This result shows the success of the *method of replays* in emulating concerted action of MANET nodes.

4.4.2 Weight Assigned to Spatial Distribution

By default, DWP assigns equal weight to the target hotspot measurement distribution and the persistence time distributions (*coordscount* and *perstime* respectively) while determining the replay with the best match, since it is not possible to know *a priori* which one will have a more significant impact on generating realistic traces.

In this evaluation, we vary the weightage assigned to the *coordscount* distribution vis-a-vis the *perstime* distribution for the campus scenario, while keeping all other parameters at their default values. As in the preceding evaluation, we save and reuse hotspot locations and node schedules to eliminate their effect on the observations. The results are shown in figure 4.13

We observe from figure 4.13 that there is no clear indication as to which value of the weighting factor generates the most accurate traces. Both, very high and very low weights tend to give more realistic results, whereas equal weights for both distributions produces less accurate traces.

4.5 Concluding Remarks

This section summarizes our observations about DWP performance across all the evaluations conducted in this chapter. The results show that there is more to realistic mobility modeling than setting realistic values for RWP pause time and speed parameters. The comparisons between real traces and DWP as opposed to RWP clearly demonstrate the relevance of incorporating spatial and temporal dependencies between node movements into the mobility model. The results also prove the ability of DWP to adapt to spatial-temporal characteristics of specific scenarios, given sample traces. During the course of the evaluation, we noted a number of strengths and limiting factors of DWP.

While DWP is not always accurate in terms of absolute performance, the performance trends indicated by DWP are in most cases far more representative of the real world than random waypoint model. More specifically, as the scenario gets less structured and the degree of freedom available to a node in selecting mobility parameters reduces, the DWP traces become more accurate, both in terms of trend and absolute values. The realism in terms of performance trends is primarily due to the characterization of scenario-specific peculiarities in spatial node distribution and the important effect it has in determining routing performance. The lack of accuracy in absolute terms may be explained by the fact that DWP makes only a best-effort attempt to reflect spatial-temporal dependencies and there is still a considerable amount of randomness in the layout as well as relative spatial separation of hotspots. The generality of DWP comes at the cost of being agnostic to the level of detail that is sometimes essential to reliably reflect absolute performance results.

Scenarios such as the university campus, where nodes are paused for a large proportion of time are particularly vulnerable to the initial layout of hotspot clusters. Once the hotspots are designated, nodes will typically pause there for long durations; hence the performance becomes sensitive to the spatial separation between hotpots. Since hotspots once designated are immutable, their initial layout plays a very important role in such scenarios. However, as we observed during the intensive evaluation, the *method of replays* proves particularly effective in such cases. When the number of replays is sufficiently large, the *method of replays* is able to reproduce the original spatial-temporal properties with a high level of accuracy inspite of being constrained by the spatial layout of hotspots. It must be noted however, that the larger number of replays entails a higher computational overhead.

The second major shortcoming is the possible fragmentation of clusters depending on the distance threshold parameter. The problem particularly manifests itself in cases where hotspot dimensions vary greatly, as it becomes difficult to fix an appropriate distance parameter for the hotspot detection algorithm.

The expectation from a hybrid mobility model is to produce results that are more indicative of real world MANET protocol performance than purely stochastic models, with less detailed information and considerably lower overhead as compared to detailed simulation paradigms. In summary we may conclude that DWP by and large fulfills these expectations.

Chapter 5

Summary and Future Work

In this thesis we propose a novel adaptive mobility model that uses real traces as a hint to impart a realistic flavor to MANET protocol simulations. We identify node movement characteristics relevant to realistic mobility modeling and formulate algorithms to measure these from provided sample traces. We also propose techniques to generate traces with similar characteristics as those measured from real data. Next, we implement a framework to streamline the entire analysis and trace generation processes. Finally, we evaluate a commonly used MANET routing protocol with traces generated by the framework for various reference scenarios. A comparison with the performance indicated while using the original traces and random waypoint model highlights the merits and shortcomings of our approach.

5.1 Future Work

Our work on the Directed Waypoint model may be extended in the following directions:

- The effect of the distance parameter used in the hotspot detection and persistence time computation algorithms on the final results needs to be systematically studied. Currently, these parameters are selected based on intuitive heuristics. A more precise and intelligent method is needed to accurately characterize hotspot information.

- Our evaluations show that the initial layout of clusters in DWP may affect the accuracy of results in certain scenarios. The current practice of designating and freezing hotspots *a priori* in the trace regeneration phase may therefore be reexamined. The idea of nomadic hotspot clusters may be worth exploring, especially for scenarios with low mobility. An alternative approach would be to measure the distribution of path durations in the real data and layout the clusters in such a way that the inter-hotspot distances closely match the measured distribution.

- The performance evaluation conducted in this thesis focuses on the overall reliability of DWP traces vis-a-vis the reference traces. An intensive evaluation that individually examines the effect of each of DWP's features is needed for an analytical study of DWP. For instance, the effect of the epoch length selected in relation to the simulation area size on the accuracy of the final traces could be studied to establish precise methods to select the epoch length.

- More rigorous metrics to measure and characterize spatial-temporal properties may be proposed and evaluated. An important requirement for any such metric is that one must be able to measure them from real data. The current metrics are simple and intuitive in nature, but can easily be measured from real data. Their drawback is that they rely on pairwise operations between nodes. Consequently, most of the trace analysis and regeneration algorithms have quadratic run-time complexity in the number of nodes.

Appendix A

Trace Analysis Algorithms

Algorithm 1 Surface Concentration, *surfconc*

Input: Sample traces of the form *(xcoords[1 . . . n][1 . . . t],ycoords[1 . . . n][1 . . . t])*

Output:

1. *nonzerocoords*, the set of all coordinate pairs which appear in the traces with their time of recording.
2. *coordscount*, the number of occurrences for each coordinate pair belonging to *nonzerocoords*.
3. *pausecount*, the number of times each coordinate pair in *nonzerocoords* appears in successive recorded time instants.

for $i = 1$ to n **do**
  for $j = 1$ to t **do**
    if $xcoords[i, j] > 0$ and $ycoords[i, j] > 0$ **then**
      increment $coordscount[xcoords[i, j],\ ycoords[i, j]]$
      add to *nonzerocoords* the tuple $[(xcoords[i, j], ycoords[i, j]), j]$
    end if
    if $xcoords[i, j] = xcoords[i, j - 1]$ and $ycoords[i, j] = ycoords[i, j - 1]$ **then**
      increment $coordscount[xcoords[i, j],\ ycoords[i, j]]$
    end if
  end for
end for

Algorithm 2 Hotspot Identification, *find-hotspots*

Input :

1. *nonzerocoords* as computed in Algorithm 1, consisting of [*loc*,*time*] tuples
2. Distance threshold, d
3. Time threshold, t

Output : *cl*, the set of clusters corresponding to hotpots in the recorded traces

for all i such that $i \in nonzerocoords$ **do**
    assign i to $cl[i]$
end for
for all i,j such that $i \in nonzerocoords$, $j \in nonzerocoords$ and $i \neq j$ **do**
    if $distance(loc_i, loc_j) < d$ and $(time_i - time_j) < t$ **then**
        merge $cl[i]$ and $cl[j]$
    end if
end for

Algorithm 3 Persistence time computation, *perstime(xcoords,ycoords)*

Input:

1. Reference traces of the form *(xcoords[1 . . . n][1 . . . t],ycoords[1 . . . n][1 . . . t])*
2. Distance threshold, d
3. *epochlength*, the duration of each epoch

Output: *grouptime*, the histograms representing the distribution of persistence times

$epochs \leftarrow round(t/epochlength) + 1$
for all i,j where i and j are nodes **do**
  $e \leftarrow 1$
  $age \leftarrow 0$
  for $clock = 1$ to t **do**
    if $mod(clock, epochlength) = 0$ **then**
      $grouptime[e + 1] \leftarrow grouptime[e]$
      $e \leftarrow e + 1$
    end if
    if $distance(i, j) < d$ **then**
      $age \leftarrow age + 1$
    else if $age > 0$ **then**
      Depending on value of age, increment appropriate interval of $grouptime[e]$
      $age \leftarrow 0$
    end if
  end for
  if $age > 0$ **then**
    Depending on value of age, increment appropriate interval of $grouptime[epochs]$
  end if
end for

Algorithm 4 Pausetime computation

```
Input: Reference traces of the form (xcoords[1 ... n][1 ... t],ycoords[1 ... n][1 ... t])

Output : pausetimes, the histogram representing the distribution of pause durations

pausetimes ← 0 for all intervals
for all nodes i do
  paused ← 0
  for j = 2 to t do
    if xcoords[i, j − 1] = xcoords[i, j] and ycoords[i, j − 1] = ycoords[i, j] and
    xcoords[i, j] > 0 and ycoords[i, j] > 0 then
      paused ← paused + 1
    else if paused > 0 then
      Depending on value of paused, increment appropriate interval of pausetimes
      paused ← 0
    end if
  end for
end for
```

Algorithm 5 Node lifetime computation

Input: Reference traces of the form *(xcoords[1 ... n][1 ... t],ycoords[1 ... n][1 ... t])*

Output : *lifetimes*, the histogram representing the distribution of node lifetimes

$lifetimes \leftarrow 0$ for all intervals
for all nodes i **do**
  $age \leftarrow 0$
  $startflag \leftarrow FALSE$
  for $j = 1$ to t **do**
    if $startflag = TRUE$ and $xcoords[i, j] > 0$ and $ycoords[i, j] > 0$ **then**
      $age \leftarrow age + 1$
    else if $startflag = TRUE$ **then**
      Depending on value of age, increment appropriate interval of $lifetimes$
      $age \leftarrow 0$
      break
    else if $xcoords[i, j] = 0$ and $xcoords[i, j + 1] > 0$ and $ycoords[i, j] = 0$ and $ycoords[i, j + 1] > 0$ **then**
      $startflag \leftarrow TRUE$
    end if
  end for
  if $age > 0$ **then**
    Depending on value of age, increment appropriate interval of $lifetimes$
  end if
end for

Algorithm 6 Bulk arrival size and inter arrival time computation

Input: Reference traces of the form *(xcoords[1 . . . n][1 . . . t],ycoords[1 . . . n][1 . . . t])*

Output :

1. $arrivalsizes$, the histogram representing the distribution of bulk arrival sizes

2. $interarrivaltimes$, the histogram representing the distribution of interarrival times

$lastarrival \leftarrow 0$

for $j \leftarrow 2$ to t **do**

  $currentsize \leftarrow 0$

  $thisinterval \leftarrow FALSE$

  for $i \leftarrow 1$ to n **do**

    if $xcoords[i, j-1] = 0$ and $xcoords[i, j] > 0$ and $ycoords[i, j-1] = 0$ and $ycoords[i, j] > 0$ **then**

      $currentsize \leftarrow currentsize + 1$

      $thisinterval \leftarrow TRUE$

    end if

  end for

  Depending on value of $currentsize$, increment appropriate interval of $arrivalsizes$

  if $thisinterval = TRUE$ **then**

    $latestinterarrival \leftarrow j - lastarrival$

    Depending on above value, increment appropriate interval of $interarrivaltimes$

    $lastarrival \leftarrow j$

  end if

end for

Appendix B

Trace Regeneration Algorithms

Algorithm 7 Replay Comparison

Input:

1. *(xcoords[1 ... n][1 ... t],ycoords[1 ... n][1 ... t])*, committed traces so far
2. *replays*, the number of replays
3. *genxcoords[replays][1 ... n][1 ... epochlength]*, traces for each replay
4. *grouptimes*, normalized target distribution of persistence times
5. *clusters*, normalized target *coordscount* distribution for each hotspot cluster
6. *normfactor*, normalization factor for the *coordscount* distribution

Output : *winner*, the replay which results in the best match with *grouptimes*

for $r \leftarrow 1$ to $replays$ **do**
  $score[r] \leftarrow 0$
  Merge $genxcoords[r]$ and $genycoords[r]$ with $xcoords$ and $ycoords$ respectively
  $gengrouptimes \leftarrow$ *perstime(xcoords,ycoords)*
  $perstimedifference \leftarrow \sum |gengrouptimes - grouptimes|$ for all epochs so far
  $gencoordscount[i] \leftarrow 0$ for all i where $i \in clusters$
  for all locations *loc(x,y)* in current replay **do**
    if $loc(x, y) \in clusters[i]$ **then**
      $gencoordscount[i] \leftarrow gencoordscount[i] + 1$
    end if
  end for
  $coordsdifference \leftarrow \sum |gencoordscount - coordscount|$ for all clusters
  $score[r] = (perstimedifference + normfactor * coordsdifference)$
end for
$winner \leftarrow minimum(score)$

Bibliography

[1] "Rfc 2501 - mobile ad hoc networking (manet): Routing protocol performance issues and evaluation considerations," January 1999.

[2] Chlamtac, M. Conti, and J. Liu, "Mobile ad hoc networking: Imperatives and challenges," *Ad Hoc Network Journal*, vol. 1N.1, January-February-March 2003.

[3] F.Bai, N.Sadagopan, and A.Helmy, "Important: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *Proceedings of IEEE INFOCOM 2003*, (San Francisco, CA), March/April 2003.

[4] P.Johansson, T.Larsson, N.Hedman, B.Mielczarek, and M.Degermark, "Scenario-based performance analysis of routing protocols for ad hoc networks.," in *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, (Seattle, Washington), 1999.

[5] Y.Xu, S.Bien, Y.Mori, J.Heidemann, and D.Estrin, "Topology control protocols to conserve energy in wireless ad hoc networks," tech. rep., University of California, Los Angeles, Center for Embedded Networked Computing, January 2003.

[6] T. Camp, J.Boleng, and V.Davies, "A survey of mobility models for ad hoc network research.," *Wireless Communications & Mobile Computing (WCMC) Special issue on Mobile Ad Hoc Networking: Research,Trends and Applications*, vol. 2(5), pp. 483–502, 2002.

[7] C.Bettstetter and C.Wagner, "The spatial node distribution of the random waypoint mobility model.," in *Proceedings of the 1st German Workshop on Mobile Ad hoc Networks (WMAN)*, (Ulm, Germany), March 2002.

[8] D. Tan, S. Zhou, J. Ho, J. Mehta, and H. Tanabe, "Design and evaluation of an individually simulated mobility model in wireless ad hoc networks," in *Communication Networks and Distributed Systems Modeling and Simulation Conference 2002*, (San Antonio, TX.), 2002.

[9] J. Jetcheva, Y. Hu, S.PalChaudhuri, A. Saha, and D. Johnson., "Design and evaluation of a metropolitan area multitier wireless ad hoc network architecture," in *Proceedings of the 5th IEEE Workshop on Mobile Computing Systems and Applications (WMCSA 2003)*, (Monterey,CA), October 2003.

[10] C.Barrett, M.Marathe, and et al., "A mobility and traffic generation framework for modeling and simulating ad hoc communication networks," tech. rep., Los Alamos National Laboratory, 2002.

[11] C.Perkins and E. Royer, "Ad-hoc on-demand distance vector routing," in *WMCSA Second 1999*, vol. 00, 1999.

[12] C.Chiang, *Wireless Network Multicasting.* PhD thesis, University of California,Los Angeles, 1998.

[13] V.Davies, "Evaluating mobility models within an ad hoc network," Master's thesis, Colorado School of Mines, 2000.

[14] X.Hong, M.Gerla, G.Pei, and C.Chiang, "A group mobility model for ad hoc wireless networks.," in *Proceedings of the ACM/IEEE MSWIN'99*, (Seattle,Washington), August 1999.

[15] D.Johnson and D.Maltz, "Dynamic source routing in ad hoc wireless networks," *Mobile Computing*, pp. 153–181, 1996.

[16] C.Bettstetter, "Mobility modeling in wireless networks: categorization, smooth movement, and border effects," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 5, no. 3, pp. 55–66, 2001.

[17] W. Navidi and T. Camp, "Stationary distributions for the random waypoint mobility model," *IEEE Transactions on Mobile Computing*, vol. 3,no. 1, pp. 99–108, January-February 2004.

[18] A.Cavilla, G.Baron, T.Hart, L.Litty, and E. de Lara, "Simplified simulation models for indoor manet evaluation are not robust," in *SECON 2004*, (Santa Clara, CA), 2004.

[19] J.Tian, J.Haehner, C.Becker, I.Stephanov, and K.Rothermel, "Graph-based mobility model for mobile ad hoc network simulation," in *Proceedings of the 35th Annual Simulation Symposium*, (San Diego, CA), April 2002.

[20] A.Jardosh, E.M.Belding-Royer, K. Almeroth, and S.Suri, "Towards realistic mobility models for mobile ad hoc networks.," in *Proceedings of the 9th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, (San Diego, California), September 2003.

[21] M.Hollick, T.Krop, J.Schmitt, H. Huth, and R.Steinmetz, "A hybrid workload model for wireless metropolitan area networks.," in *Proceedings of VTC'03*, (Orlando, Florida), October 2003.

[22] M. Musolesi, S. Hailes, and C. Mascolo, "An ad hoc mobility model founded on social network theory," in *Proceedings of the 7th ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM 2004)*, (Venezia, Italy), October 2004.

[23] J.Cano, P.Manzoni, and M.Sanchez, "Evaluating the impact of group mobility on the performance of mobile ad hoc networks," in *Proceedings of IEEE ICC 2004*, (Paris, France), June 2004.

[24] W.Hsu, K.Merchant, H.Shu, and A. Helmy, "Weighted waypoint mobility model and its impact on ad hoc networks."

[25] A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilit, "Place lab: Device positioning using radio beacons in the wild," in *Proceedings of Pervasive 2005*, (Munich, Germany), 2005.

[26] "About gps website http://www.garmin.com/aboutgps."

[27] J.Yoon, M. Liu, and B. Noble, "Random waypoint considered harmful," in *Proceedings of INFOCOM 2003*, April 2003.

[28] J.Kang, W.Welbourne, B.Stewart, and G.Borriello, "Extracting places from traces of locations," in *Proceedings of the 2nd ACM international workshop on Wireless mobile applications and services on WLAN hotspots*, (Philadelphia,PA), 2004.

[29] S.McCanne, S.Floyd, K.Fall, and K. et al., "Network simulator - ns2," 1997.

[30] "North carolina state univerisity tracs schedule http://www2.acs.ncsu.edu/regrecords."