MOBILEMAN

IST-2001-38113

Mobile Metropolitan Ad hoc Networks

MOBILEMAN

Economic value of self-organisation paradigm and market access

Deliverable D18

Contractual Report Preparation Date: 31 August 2005
Actual Date of Delivery: 31 October 2005 (*)
Estimated Person Months: 16
Number of pages: 76

Contributing Partners: Consiglio Nazionale delle Ricerche (Italy), University of Cambridge (UK)

Authors: Raffaele Bruno, Marco Conti (CNR), Jon Crowcroft, Wenjun Hu, Elgan Huang, Meng How Lim, Andrea Passarella (Cambridge)

(*) new schedule accepted by the Project Officer

Abstract: this deliverable discusses market opportunities offered by the use of MANET technologies. We firstly focus on traditional MANETs, and present usage scenarios where MANETs represent either a cost-effective way of re-engineering existing applications, or the technological enabler for brand-new services. Then, we consider two possible evolutions of the MANET paradigm (namely, mesh and opportunistic networking) in the short/medium, and long time frame, respectively. We analyse these scenarios, and highlight the further potentialities that they can offer.

Project funded by the European Community under the "Information Society Technologies" Programme (1998-2002)

Summary

This document explores the potentialities of Mobile Metropolitan Ad hoc Networks from an economic point of view. Therefore, together with Deliverables D15 and D19 it represents the main outcome of the Workpackage 5.

In this document we present several scenarios in which the use of MANET technologies is the basis for the development of services valuable for end users, and a cost-effective solution for the service providers. We give examples of how re-engineering existing applications based on the MANET paradigm allows to i) reduce the maintenance costs for the service providers, ii) reduce the entry barriers for new competitors, and thus iii) reduce the costs for end users. Moreover, we present scenarios in which MANETs allow to develop brand-new applications and services. Finally, we discuss further potentialities that arise by extending the MANET paradigm towards two main directions, i.e., *mesh networks* and *opportunistic networking*.

For clarity of presentation, this document is made up of three sections. Section 1 analyses the *legacy* MANET networking scenario, i.e., a scenario where the MANET is purely infrastructure-less, and no pre-existing infrastructure is used to build the mobile network. Several usage scenarios are presented, and evaluated from a technical/economic point of view. The approach taken in this analysis is original. The main research approach used by the MANET research community has been a bottom-up one. Lots of protocols and networking solutions have been proposed to deal with the challenging networking environment, in order to support legacy applications also in MANETs. In this view, MANETs are not seen as an opportunity for applications, but as quite hostile environment for legacy applications to operate in. On the contrary, we envision *novel* application scenarios that *leverage* the key MANET features in order to provide value to the end user. The required networking solutions come in a second stage, in order to address the real needs of these scenarios. Such a top-down approach has been often neglected by the research community, while we believe that it represents a key avenue to successfully bring MANET technologies into the market. Finally, since a key element for evaluating the proposed scenarios is the user mobility pattern, Section 1 also presents realistic models of user mobility.

Ever more widespread consensus is diffusing in the research community on the fact that foreseeing very large, flat, legacy MANETs is not realistic. Furthermore, interconnecting MANET islands with the legacy Internet is not that easy. On one hand, some MANET nodes should be also connected to the *wired* Internet (or to an 802.11 Access Point), and act as gateways between the MANET and the Internet worlds. This approach requires the presence of *costly* Internet infrastructure close to the MANET, which is quite a limiting factor. On the other hand, some MANET nodes should be also connected to 2.5G/3G cellular networks, which would act as a bridge between the MANET and the legacy Internet. Though this solution requires less fixed infrastructure, it is very costly, and is not able to provide an acceptable bandwidth.

We believe that a very promising direction to address both the MANET scalability and interconnection issues in a cost-effective way is represented by *Mesh Networks*. Therefore, Section 2 is devoted to analysing the viability of this approach. Mesh Networks are multi-tier networks generally composed by three tiers. The first tier is represented by MANETs. Some nodes in each MANET are (wirelessly) connected to Wireless Routers, which represent the second tier. Specifically, Wireless Routers are static nodes (usually with more capabilities than MANET nodes) that form a mesh by establishing wireless links with each other. Finally, some Wireless Routers are connected to Access Points, which provide Internet access to the whole Mesh Network. Even though the Mesh paradigm includes some infrastructure (the Wireless Router tier), it should be pointed out that this infrastructure is completely wireless, and can be built on top of open standards such as 802.11 or 802.16. Furthermore, Mesh Networks can deliver reasonable bandwidth to the end users. Therefore Mesh Networks are

quite a cheaper and more effective solution than the ones mentioned before. In the short/medium term, we envision this paradigm to be the most promising evolution of legacy MANETs.

Section 2 presents usage scenarios for Mesh Networks, and surveys both proprietary and open solutions that are being developed in the market. Examples are systems for intelligent public transportation and public safety, and system providing Internet access to rural and scarcely populated areas. Moreover, this section presents the main efforts that are being carried out by the International Standardisation Bodies to provide open standards explicitly conceived for Mesh Networks, and, finally, discusses the main research challenges that are still open.

Finally, section 3 focuses on MANET evolution in a longer time frame. We envision *opportunistic networking* as one of the most intriguing scenarios from this standpoint. In such scenario, each device forwards data in an opportunistic way, i.e., by exploiting any possible contact with other devices. For example, a contact opportunity is represented by two people walking in the same corridor. Their buetooth/wifi enabled mobile phones get in touch and forward data to each other, “hoping” that the other device will carry the information closer to the eventual destination. Clearly, this scenario opens very challenging research directions, and paves the way for new applications, viable from an economic point of view. For example, due to the ever more widespread diffusion of mobile devices, the infrastructure costs of applications based on opportunistic networking could be negligible, if not eliminated at all. This document includes experimental results that represent a bottom-line to design and evaluate opportunistic networking solutions. Opportunistic networking may be envisioned as a building block of future, heterogeneous, networks. The user will be able to seamlessly switch between different networking technologies, and will dynamically choose the one that better suits her needs. In other words, we argue that future networking environments will be heterogeneous not only in the network technologies, but also in the *user-QoS needs*. For example, depending on the type of application and message content, the same user could require either costly, high-reliable, low-latency delivery, or cheap, best-effort delivery. This argument calls for *suites* of networking solutions to be included in mobile networks, rather than single, one-fit-all, protocols. To complement the routing schemes envisioned during the project, we thus present a geo-routing approach that is able to exploit even imprecise node-location information.

Table of Contents

1. LEGACY MULTI-HOP AD HOC NETWORKS

1.1. Introduction

In this section we elaborate on two key aspects for deploying MANETs in the real world, specifically i) realistic usage scenarios, and ii) realistic mobility models. The networking scenario we take into consideration is a legacy MANET, i.e., a MANET which does not rely on any pre-existing networking infrastructure.

In previous deliverables we provided a conceptual framework for evaluating MobileMAN business models, and proposed usage scenarios whose value chains have been analysed through this framework. In the first part of this section (namely, Sections 1.2 and 1.3) we briefly recall these usage scenarios, and deeply analyse the feasibility of one of them, i.e., the city taxi scenario. This complements the business analysis of the value chains presented in previous deliverables. All in all, we show that MANETs can be a valuable way of improving current radio dispatch systems for cab companies. Specifically, re-engineering the radio dispatch system through MANETs allows to i) reduce the cab company recurring costs related to the radio licensing, since MANETs use unlicensed spectrum; ii) reduce the entry barrier for new competitors, since the MANET radio dispatch system is based on open standards; and thus iii) reduce the final cost for cab users, since more competitors can be expected to operate in the market.

While analysing MANETs from an economic/technical standpoint, we realised how important is adopting realistic models of user mobility. Unfortunately, though well suited for analytical purposes, mobility models traditionally used in the literature (e.g., the random waypoint model) poorly reflect real-user movement patterns. Therefore, we propose the MIRRORS framework, which allows us to build mobility models inspired by realistic scenarios, such as users walking along streets, or driving a car. Specifically, we show how the city taxi scenario can be analysed within the MIRRORS framework.

1.2. Usage Scenarios

In previous deliverables, we described the possible real world deployments of MobileMAN for several situations including the Shopping Mall, and the City Cab Taxi Ad Hoc dispatch network. Here we look at two further possibilities, based on extending the model of deployment to hybrid networks, in both directions: where there may be occasions with more connectivity, and where there may be less. We also assume that as devices become cheaper and radio integration more advanced, location data may become a first class piece of the architecture, although its accuracy may be in question in many situations:

1.2.1. Personalized Radio on the Road

Many users have portable devices (e.g. iPod etc) with large amounts of storage. Many users also have wireless laptops, and also have computers at home with peer-to-peer applications. We can combine these into an interesting new application: personalized radio on the road. The idea is to provide peer-to-peer file sharing amongst vehicles on the highway – given enough cars with users carrying 802.11 capable devices, the amount of content potentially is quite large – a search for a particular reasonably popular song is likely to yield success in a short time. We have looked at extending the MobileMAN P2P application over this scenario. We have some initial studies conducted in another project with Intel, of 802.11 between vehicles travelling at 100Km/h (and Infocom paper from Bremen cites similar results), one can transfer easily 10Mbytes of data during a contact opportunity.

1.2.2. Mixed Reality Games

A second scenario could be based on the UK Equator project's Mixed Reality Game "Can you see me now?" (see earlier MobileMAN interactions with them). This relies on location data, albeit, not accurately. Essentially the game is a mix of three traditional games “hide and seek”, “treasure hunt” and “blind man’s buff”: Players have hand held computers that display a map with the x,y,z position of clues as to how to find the treasure marked on an overlay of the real world, and illustrate the approximate position of this, and all other players. The object of the game is to go to the location of the clues in the real world, and then use them to find the treasure. If other players see and catch a player, the player is “out”.

Role playing versions of such a game (viz Everquest etc) could easily be envisaged. There is a lot of commercial interest in such games.

Clearly, the game could be played using infrastructural networks where available (e.g. GSM/GPRS) but more accurate location data is needed too, although not too accurate (10’s of meters would be fine). To understand whether the applications above can be made to work, we need to understand not just the technical aspects of the networks, but also the economic impact of the technical shortcomings of the early deployment.

1.2.3. City Taxi Scenario

In recent years there has been a growing interest in using mobile ad hoc networks (MANETs) for both civilian and commercial purposes. Within the automotive industry there are already a number of efforts geared towards utilizing communications technology to improve automotive safety, provide passengers with information and entertainment, and achieve smooth traffic flow on the roads [1]. However, there has been little effort to tie business and technical aspects together in order to design solutions that are both technically and financially feasible; most research is currently focused primarily on detailed technical issues. Standard scenarios commonly quoted for MANETs tend to be either unnecessary or too limited in scope.

In this section, we propose a novel application scenario in the form of a MANET dispatch system for taxis based on [2] and pragmatically evaluate the system’s feasibility from both financial and technical viewpoints.

In the following sections, we begin by laying out the vision for the dispatch system in detail, analyzing the value chain and value proposition for each player and evaluating the operating costs of existing dispatch technologies. This is similar to the tussle considerations in [26]. Applications for MANETs need to bring value to users, whilst at the same time provide business incentives for corporate participation. We then investigate the technical feasibility of the system, using a realistic mobility and propagation model to simulate the system’s real world nature and scale. We analyse various factors and limitations that affect system performance and look into methods of improving desired performance metrics.

Despite the section’s focus on the application proposed, the main objective of this work is to highlight general principles and design considerations that are generally applicable to large-scale highly-mobile MANET applications operating in micro cell environments (mostly vehicular).

1.3. The City Taxi Scenario in Detail

One of the most important components of a city taxi company’s operations is its dispatch unit, which informs individual taxis about passenger pickups, assigns passengers to empty taxis, and passes on additional information such as directions and news about weather or traffic conditions.

Traditionally, taxi companies have relied on radio dispatchers for these tasks. Radio dispatch systems have variable QoS issues, and there are significant costs associated with installing, running and maintaining them. In many cases, radio licenses need to be obtained to operate these systems, adding considerable cost and creating barriers to entry for small companies. Economies of scale mean that large companies are able to share the cost of radio dispatch amongst many drivers, whilst smaller players are unable to afford these systems.

However, the popularisation of mobile phones now means that smaller companies are able to enter the market by relying on public mobile networks as their dispatch systems. Unfortunately, this method has many drawbacks including a high average cost per call, and limited communication between drivers. Mobile phone calls are harder to coordinate and less time efficient; dispatchers can only communicate with one driver at a time. Voice relayed instructions are also subject to higher error rates.

Many bigger companies have in turn upgraded their own systems to include smart software with GPS which allows tracking of taxis and efficient allocation of jobs, improving productivity and utilisation of resources. However, the hardware and software required is costly (costing over £630,000 for a company with 300 taxis [3]) and its operation involves transferring data wirelessly over public or private mobile radio networks, which is a significant recurring cost.

In our scenario, a MANET based dispatch system is used for communication. Each taxi is fitted with an ad hoc device and a central ad hoc server is located at dispatch headquarters. This is a hybrid scenario in that although the system is a fully mobile ad hoc network, there is a degree of centralisation to it, with all dispatch information originating from one point, requiring all nodes to connect to it directly or indirectly to receive jobs.

A customer would call in to the taxi company's dispatch headquarters. The call would then be picked up by a dispatch handler who would enter the job details into the computer system. An automated request would then be sent via the ad hoc network to all free taxis in the vicinity. In the case of ad hoc devices enabled with location information, the nearest free taxi to the pickup point could be automatically located and assigned the job, thereby decreasing waiting times for the customer and improving turnover rates for the taxi company. The ad hoc system would constantly be updating itself automatically with relevant job and road condition information for drivers, and drivers would also be able to communicate amongst them, empowering all parties concerned with valuable information.

Taxi stands throughout the city could also be equipped with ad hoc terminals that automatically route taxis to appropriate stands when requested. Should ad hoc devices become ubiquitous, it would even become possible for customers to simply order a taxi via their own standard ad hoc devices wherever they were and have the request sent directly to the nearest taxi. The taxi would then register the pickup and send the information through the network to update other taxi drivers and headquarters of its status. With the addition of an access point at dispatch headquarters, customers would also be able to book taxis over the Internet and have the request relayed immediately via the ad hoc network.

In all cases there would be a two way flow of information; once a particular taxi driver has accepted a job, details such as the taxi's registration number, its current location and the estimated time of arrival can be sent back via the ad hoc network to the user's phone via SMS, to the taxi kiosk, to the customer's ad hoc device or to the messaging service of the Internet customer respectively, reducing any potential misunderstandings that are common with voice relayed radio dispatching services and allowing better planning on the customer's part.

If the network were secure enough, credit card payment options could also be implemented using the ad hoc terminals, with encrypted verification information sent to and from the central node.

We envision the rollout of the ad hoc dispatch system in three distinct stages. In the first stage, a medium sized company is able to deploy a system as described in this section, with stationary nodes that double as taxi stands to improve relay performance. In the second stage, multiple companies use the system and compatible standards are such that different companies' ad hoc networks can relay data across each other, with information protected by encryption. In the final stage, mobile ad hoc devices have become ubiquitous and data can be relayed through a multitude of devices.

The value chain shown in Figure 1.1 is a map of the primary players within the taxi company radio dispatch industry. These players make contracts between each other to conduct exchanges of information, money, services or a combination of the three. Dotted lines indicate the flow of information or services, for example the information that is shared between taxi drivers and dispatchers, whereas solid lines indicate a monetary exchange, for example when a customer pays a taxi driver for the taxi ride.

Customers will now be able to make bookings with greater ease, convenience and efficiency through a variety of different media besides the phone. Also, information flow is now two-way and customers can use the information provided for better planning. With location specific job assignments and less information congestion on the dispatch side, taxis can get to customers much more quickly, reducing their overall waiting times. The chances of getting a taxi should also be improved due to the higher turnover rate and accuracy of the system. Finally, customers will be able to enjoy a quieter environment in taxis, without constant radio chatter.

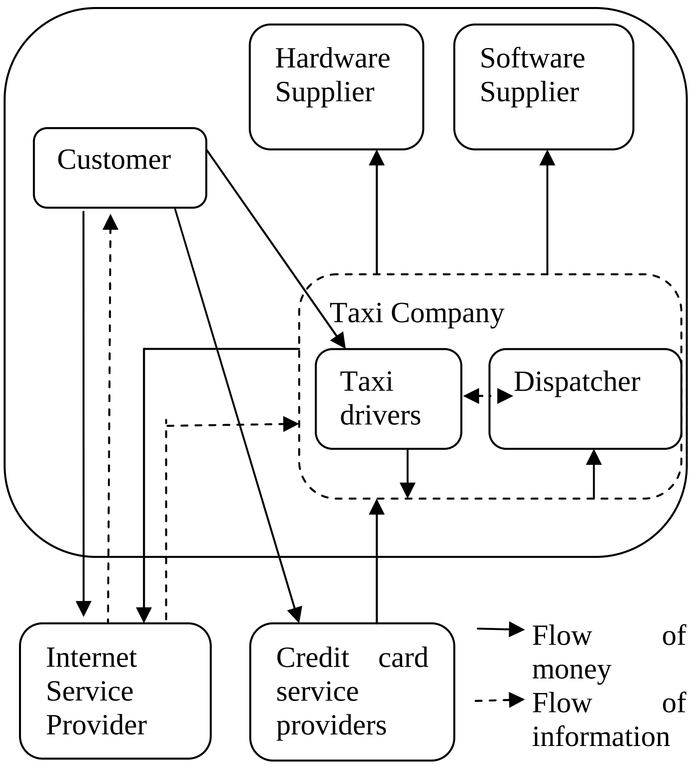


Figure 1.1. Value chain for a taxi company operation's primary players

Taxi drivers also benefit, since faster and more efficient processing of jobs potentially means faster turnover rates and less empty cruising time, thereby increasing revenue. Furthermore, decreasing reliance on voice dispatching results in fewer misunderstandings between dispatchers and drivers. With location information and a hidden alarm switch, drivers in distress can immediately be located and helped, in the event of an accident, a mugging, a carjacking etc [4]. Finally, taxi drivers can also enjoy a quieter environment in taxis.

Taxi companies benefit in many ways. Streamlining the order taking, order processing and dispatching process result in significant capacity increases, improving both revenue streams and service levels while lowering costs. The system itself could bring about significant long term savings as operational costs are marginal compared to conventional methods. High customer and employee satisfaction could improve the taxi company’s reputation and hence demand for its services. With faster information flow, management’s decision-making capabilities could be enhanced, and the reaction time for a decision to filter down to the driver level will be decreased. Location information allows taxi tracking in the case of emergencies, as well as facilitating better route forecasting and resource allocation.

Additionally, ad hoc systems are more robust than centralised systems, and network service is not dependent on service providers or subject to damaged infrastructure. Using decentralised ad hoc devices may also change the market structure by making it possible to offer off-the-shelf solutions to taxi companies, rather than forcing them to sign up with a service provider. Competition may then force service providers to offer more competitive prices

Finally, as customers get used to alternatives to booking over the phone, the workload for taxi dispatchers will decrease. Reliable automated systems will improve the accuracy and speed with which orders are taken and passed on.

1.3.1. Comparison of costs

Running a proprietary private radio network is very expensive, involving costs for infrastructure, spectrum licensing, maintenance and devices. Consequently, this option is out of the question for most taxi companies.

The alternative for many smaller companies is the use of mobile phones. However, the cost of phone calls can mount up quickly. If each driver were to spend just one minute on each call with dispatch, and make 40 of these calls per day [5], at an average cost of 20p per minute (taking into account peak and non peak calls), it would cost £2,920 per taxi annually. A medium sized company with 300 taxis would therefore be paying £876,000 pounds a year on phone bills alone. Note that such a company could move to a system with better economies of scale, but the figure is useful for comparison purposes.

Subscribing to a public network radio service is generally a much cheaper option. A typical service would be a fixed charge, push-to-talk walkie-talkie type service, as provided by Dolphin Telecommunications [6] in the UK. With a monthly fixed charge of £25 per user plus VAT, a company with 300 taxis would need to spend £105,750 annually. Although much cheaper than using mobile phones, this is still a large recurring cost and is on top of hardware costs.

Another alternative is to use a Private Mobile Radio (PMR) system which can be operated by the taxi company or outsourced to a specialist service provider. In either case a license is required from Ofcom (Office of Communications) covering the intended service area. Through the use of a single radio communications mast a range of 15-20 miles can typically be covered at a relatively low cost. However this depends on the exact service used and features required. Panther Taxis [5] uses a service from Auriga Communications which costs £3,090 a year to operate (communication only), whereas a company like Cordic [7] might charge up to £124,800 a year (£8 per week per PDA rental).

In comparison, ad hoc systems cost nothing to run above the cost and maintenance of the hardware itself. In addition, both mobile phones and the push-to-talk service are voice only communication methods (data services would cost an additional premium), whereas an ad hoc system would be based on data communications and potentially also provide location information. According to [5], 95% of bookings are made within a 10-mile radius, which makes ad hoc networks feasible and potentially more scalable than licensed services.

MANETs are also likely to be much faster to roll out and cheaper in terms of hardware costs since they do not require any installation costs or infrastructure beyond the ad hoc units themselves.

1.3.2. Technical Feasibility of City Cab Scenario

Having laid out the business case for this conceptual system, we now turn to the technical feasibility. By nature, an ad hoc system's coverage depends on a suitable spread of nodes around the required area. This means that to work properly, there must first be a sufficient number of taxis to cover the service area. In addition, the mobility of nodes poses the danger of occasional service gaps when no taxis are acting as relays within a certain area. Although these disruptions may only be momentary in nature, the unpredictability of the QoS is a risk to taxi companies who rely on the network to serve customers.

In order to investigate these and various other problems, we designed and implemented a simulation consisting of a realistic mobility and propagation model and used it to assess various metrics of performance for the radio dispatch system.

Based on the model, we developed a statistical simulation to calculate the positions at 1 second granularity, from which connectivity information is derived

The control parameters for the simulation were as follows: We used a Manhattan grid of 5km x 5km, with a uniform block size of 100m x 100m. (This layout is chosen to normalise the effects of street plan. Maybe estimate the effects of one-way streets – I might have some statistics from other simulations, so remember to ask me in a few days! Then something like: We do not consider journeys into and out of suburbs) There is one central dispatch point located at the exact centre of the grid and this doubles as a taxi stand. Each simulation run lasted for 3 hours (10,800 seconds) in simulation time after an initial warm-up period of 1,000 seconds. We chose 3 hours to give each taxi enough time to make several journeys for the speed and destination distributions to converge. The positions and connectivity information is calculated at 1 second granularity – delete. Within the grid we had a relatively conservative total of 300 ad hoc enabled taxis, therefore on average each taxi had to cover over 83,000m2. A larger taxi company would obviously have more cars, and the number of nodes would also increase if different companies shared the same ad hoc system (With 330 cars, Panther Taxis is only one of many companies in Cambridge, which has a central area of roughly 9km2). Finally, a node is considered connected only if it is reachable from the dispatch point (either directly or via a number of intermediate hops), for three consecutive seconds or more, to ensure sufficient time for connection establishment and data transfer. Since the distribution of such popular destinations can affect the results significantly, we consider the case of a small town, in which the bus/train stations are co-located. As to the uniform distribution over the remaining destinations, again, for a small town, areas out of centre are fairly equally 'favoured'.

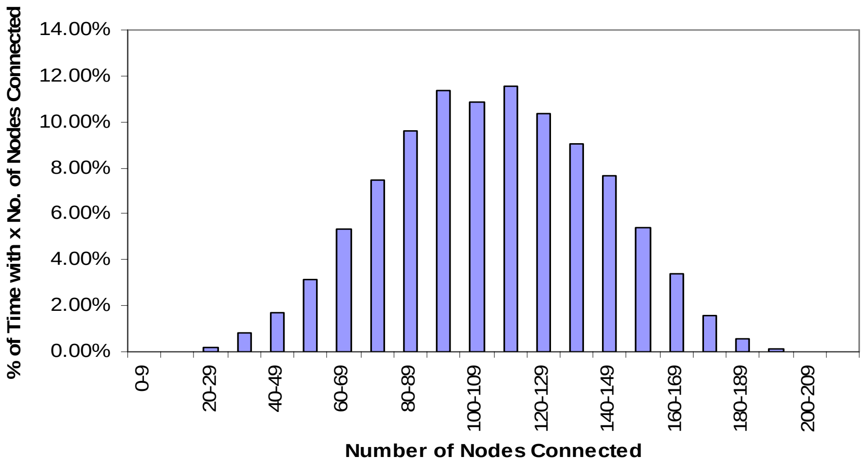


Figure 1.2 Typical Coverage Under Control Conditions

The following data were generated from up to 100 simulation runs each to ensure statistical significance. We define connectivity coverage as the percentage of taxis reachable from the central dispatch, and outage time as a continuous period of time when the taxi is not reachable from the dispatch. Figure 1.2 shows a typical graph of coverage under control conditions. Clearly, the percentage of cars connected at any given point in time fluctuates, however there are always cars within coverage and even with only 300 nodes, the mean coverage over 100 runs is 107.7 cars or 35.91% (with a 95% confidence level of 0.12%). The median and the standard deviation of the mean is 35.84% and 0.6% respectively. The distribution of the coverage is fairly Gaussian. Given the small error bounds, we omit error bars in the plots.

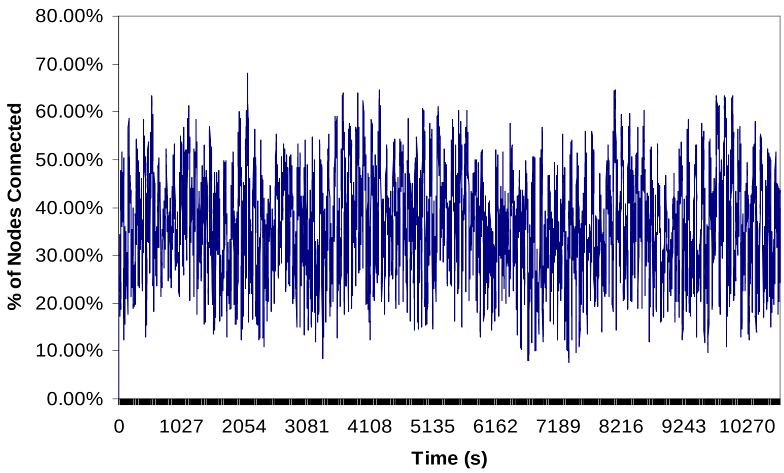


Figure 1.3: Typical Distribution of Coverage

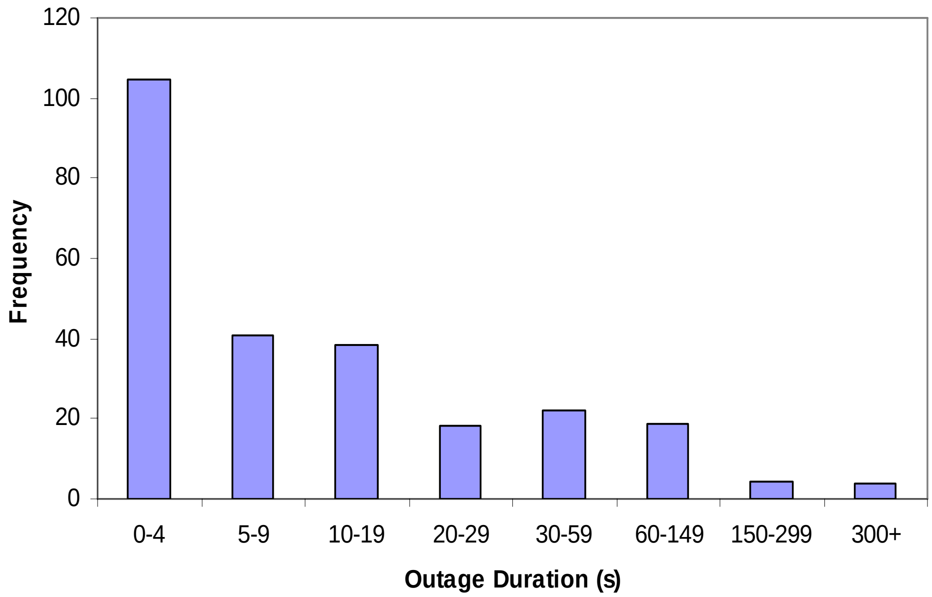


Figure 1.4: Typical Distribution of Outage Durations

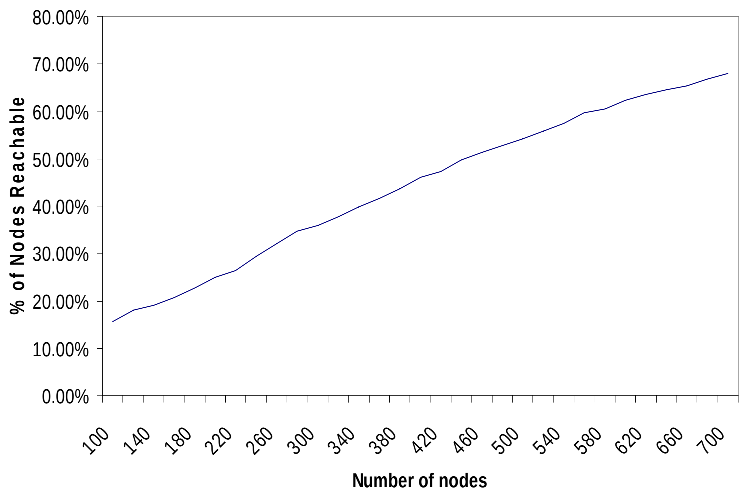


Figure 1.5: Percentage of Nodes Reachable as a Function of the No. of Nodes

We calculated the length of outage durations to assess the extent of intermittent connectivity that the system needs to be able to handle. For control conditions, the average outage time was 28.47s, with a 95% confidence level of 0.15s, which should be acceptable where real time communication is not essential. However, given the complex relative motion between nodes, the maximum time that a node might be unreachable was considerably longer, averaging 696.49s (11 min), with a 95% confidence level of 3.79s. The longest time ever observed for a node being out of coverage was 2,785s (46 min), albeit an exceptional occurrence. As shown in Figure 1.4, the majority of outages are fairly short. We also need to consider the fact that with each hop across nodes there is an associated delay that could further increase the latency. However, this is unlikely to be significant compared to the outage times of an unreachable node.

Therefore, the simulation seems to suggest that with 300 nodes in operation, real time communication is not feasible with only one third of taxis reachable at any given instant. Considering the absolute numbers of contactable taxis, however, the performance is for the purposes of the dispatch system. Also, given the short time that each node spends out of coverage on average (28.47s), all taxis should be contactable within reasonable intervals. Conversely, on occasion certain nodes will be unreachable for a longer period (for example when taking passengers to far away destinations) and backup systems may be necessary.

Intuitively, the coverage performance depends on the simulation parameters, the most obvious being the node density. Therefore, we investigated this relationship by varying – delete. We varied the number of nodes from 100 to 700. Figure 1.5 shows the percentage of nodes reachable as a function of the number of nodes.

As expected, coverage improves as more nodes are available to act as relay points. With 700 nodes, almost 70% of taxis are reachable at any given time. By increasing the number of nodes, a larger taxi company or a number of small/medium companies that use compatible systems would be able to achieve proportionally good coverage. Figure 1.6 shows the distribution of coverage for 700 nodes. Compared to Figure 1.3, more nodes are connected more frequently resulting in the distribution skewing significantly.

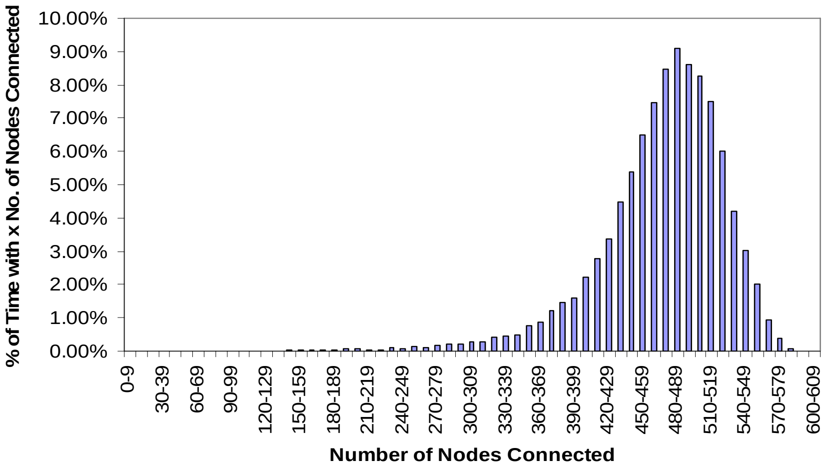


Figure 1.6: Typical Distribution of Coverage with 700 nodes

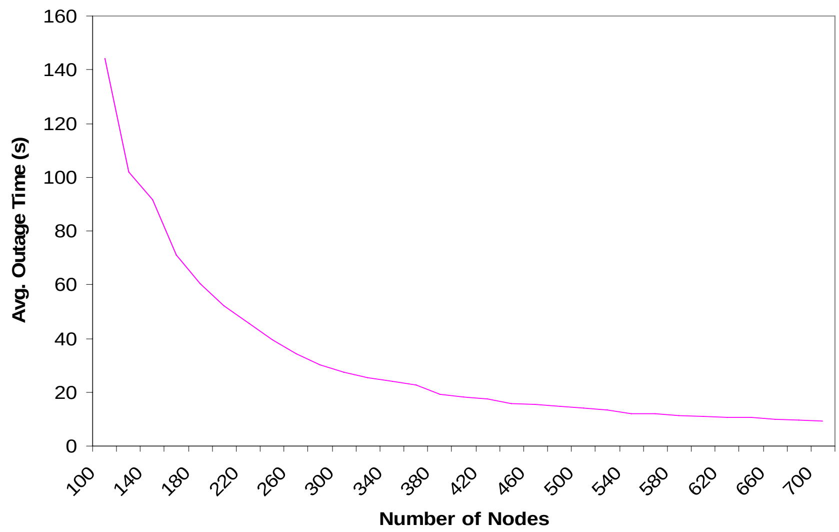


Figure 1.7: Average Outage Time as Function of No. of Nodes

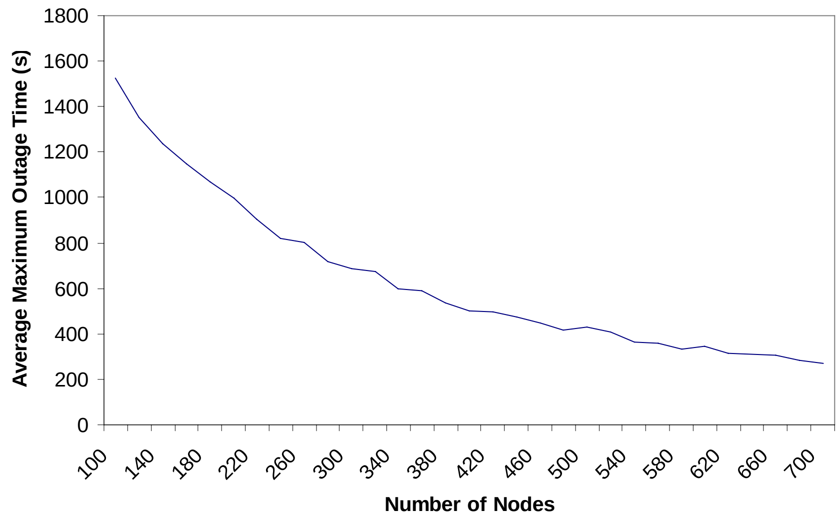


Figure 1.8: Average Maximum Outage Time as a Function of the No. of Nodes

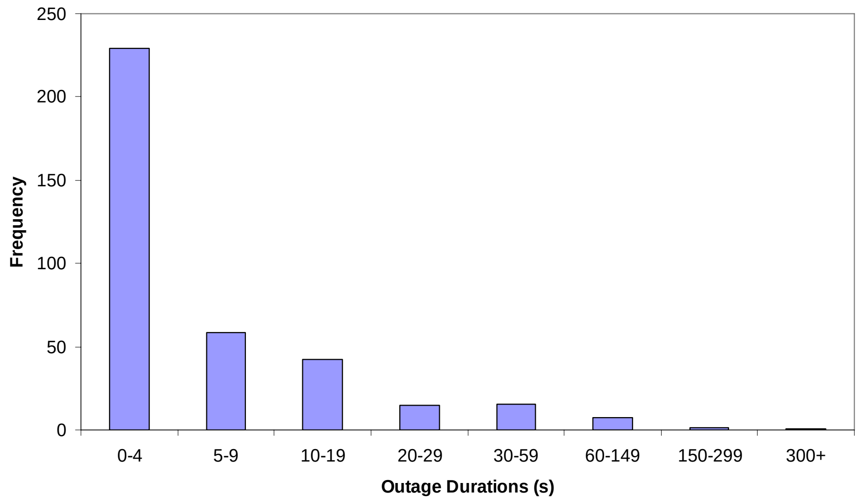


Figure 1.9: Typical Distribution of Outage Durations with 700 nodes

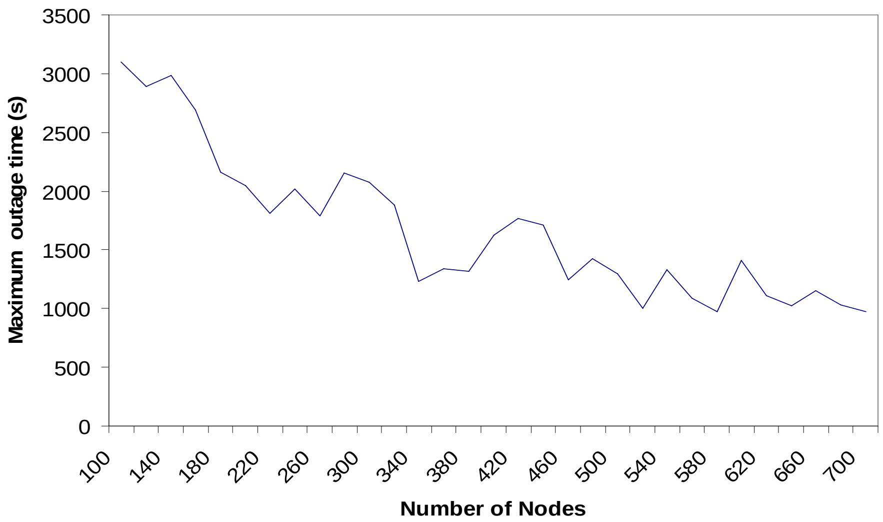


Figure 1.10: Maximum outage time

Figure 1.7 and Figure 1.8 illustrate the average outage times and average maximum outage times varying with node density. The 'average maximum outage time' is the average of all the maximum outage times of all the nodes over each simulation run. Again, the trend shows clearly that as the number of nodes rises the average outage time decreases. However, the rate of decrease grows less pronounced, such that the incremental benefit lessens, thus exhibiting saturation effects. With 700 nodes, the average time a node spends out of coverage is only 9.4s. Likewise, the average maximum time that a node spends

out of coverage decreases in the same manner, with 270.1s outage time at 700 nodes (compared to almost 700s with 300 nodes).

Figure 1.9 shows the distribution of outage durations, and once again in comparison to Figure 1.4, with 700 nodes we observe a skew to the left, with most outages being of a very short duration.

Figure 1.10 shows the maximum time that a node was observed out of coverage as a function of the number of nodes. Due to the complex movement of the nodes, there is a high degree of randomness to this, and the figures vary considerably between simulation runs.

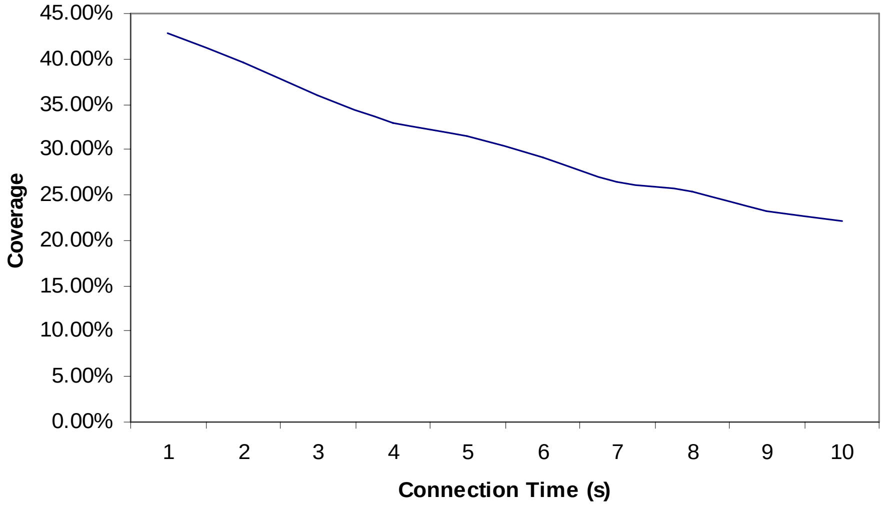


Figure 1.11: Coverage Variation with Connection Time

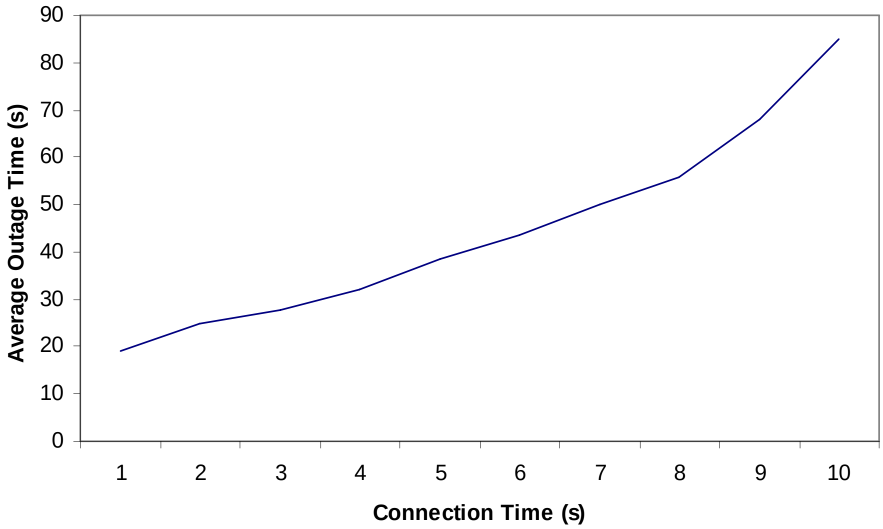


Figure 1.12: Average Outage Time vs. Connection Time

Overall, the decreasing trend is still clear, with an overall maximum outage time of 974s with 700 nodes, compared with well over 2,000s for 300 nodes. Figure 1.7 to Figure 1.10 show that as the number of nodes increases, the time that a node is likely to spend out of coverage decreases significantly and the system grows closer to being able to support near real time communications. The dispatch system would be able to operate in such conditions with users largely unaware of the delay in receiving messages. However, longer outages could still occur, necessitating backup systems.

For the ubiquitous stage, results show that with 3,000 nodes, an average coverage of 95% can be achieved.

Another parameter that greatly affects the coverage statistics is the connection time, i.e. the time to set up links and transmit data between neighbouring nodes. The default connection time is set to 3s, however this depends on the protocols and hardware. Keeping the other control parameters constant, we varied the connection time between 1 and 10s, with the results shown in Figure 1.11, Figure 1.12 and Figure 1.13.

With a connection time of 1s, 43% of nodes are within coverage on average, compared with 36% for a connection time of 3s. Given the high relative mobility of the nodes, it is intuitive that the faster a connection is established and data transferred, the more likely it is that a passing node can stay within coverage for the time required. Likewise, it becomes less likely that a node is out of coverage for a long period of time.

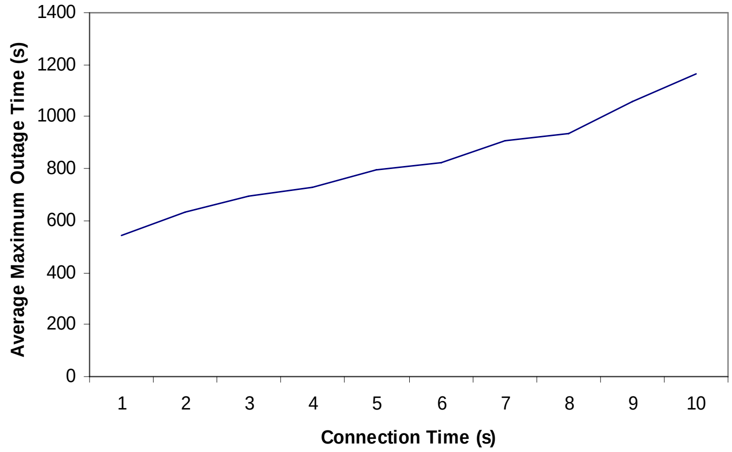


Figure 1.13: Average Maximum Outage Time vs. Connection Time

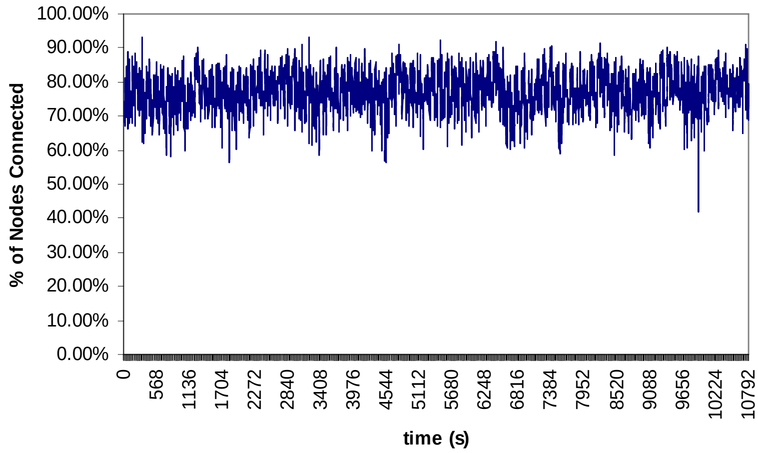


Figure 1.14: Coverage with 700 Nodes and a 1s Connection Time

By both decreasing the required connection time and increasing the number of nodes, system performance can be improved significantly. The coverage and outage results for this scenario are presented in Figure 1.14, Figure 1.15 and Figure 1.16. With 700 nodes and a 1s connection time the average coverage achieved was 77%, and the average outage time dropped to 8.8s. The majority of nodes are now connected most of the time, and generally experience very short outage durations.

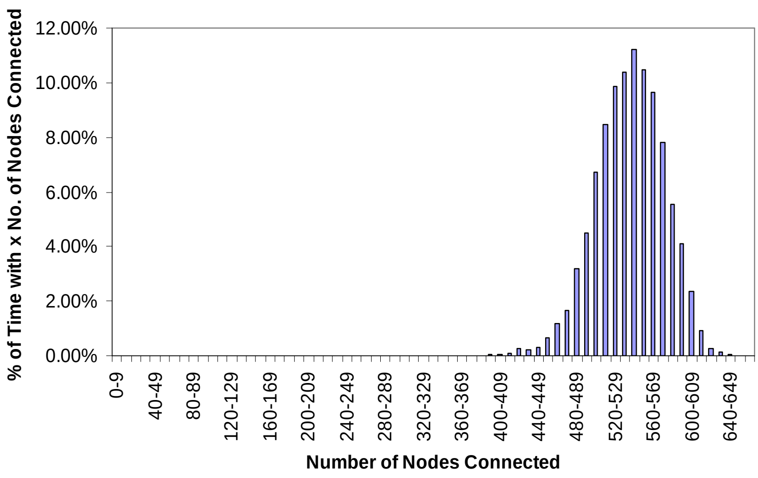


Figure 1.15: Coverage with 700 Nodes and a 1s Connection Time

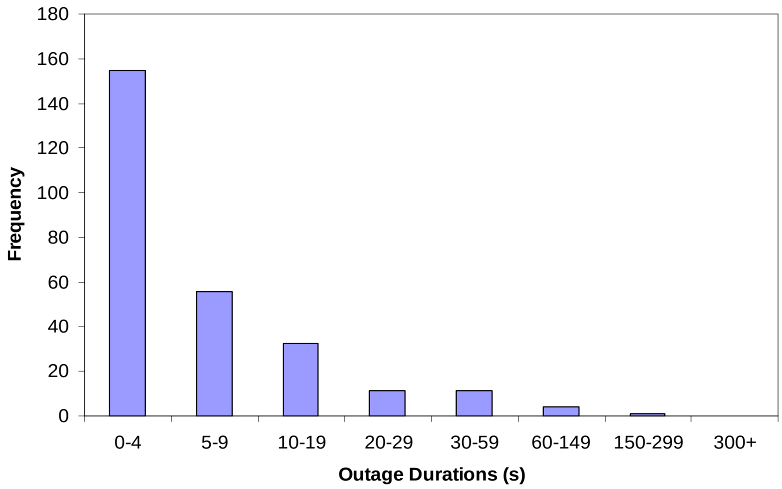


Figure 1.16: Typical Distribution of Outage Durations with 700 Nodes and a 1s Connection Time

In order to achieve better coverage for such applications it may therefore be advisable to employ proactive routing protocols, despite the increased overhead. Such routing algorithms maintain an overview of the network topology and nodes are therefore better able to establish rapid connections [21]. Other - delete techniques may also be exploited to speed up connection times, such as making use of location awareness to predict handovers.

Finally, we investigated how traffic congestion would affect the results. To be feasible, the dispatch system has to be able to operate at all hours of the day regardless of traffic conditions. Traffic is introduced by inserting cars that behave the same as taxis, except that they do not have preferred destinations and do not pause between movement sequences. We therefore varied the number of cars from 300 (only taxis) up to 30,000 (300 taxis and 29,700 regular vehicles) to study how congestion affects node distribution, mobility and coverage. If all 30,000 cars were evenly distributed throughout the map, the average distance between vehicles would be 34m. The results are shown in Figure 1.17 to Figure 1.24.

Our first observation was that the instantaneous variation in coverage was more noticeable with added congestion. Comparing Figure 1.17 with Figure 1.2, the peak to peak variations are larger, and the peaks and nulls in Figure 1.17 appear to be more regular and evenly spaced. Comparing Figure 1.18 with Figure 1.3 (30,000 cars vs. 300 cars), the distribution with congestion appears much more spread out, exhibiting bimodal characteristics. Although the mode and average coverage increase only slightly (38.4% vs. 36.3%), at any given instant, it is now much more likely that a higher number (over 140) of cars will be connected. On the other hand, it is also more likely that a lower number (below 70) are connected.

Indeed, in Figure 1.20 we can see that the standard deviation of coverage increases with congestion. However, after a steep rise it appears to flatten out at approximately 15,000 cars. Variability extends to outage times as well. Despite a nominal mean outage time change from 28.3s to 30.8s, a comparison of Figure 1.19 and Figure 1.4 shows that the distribution has spread out, making it is more likely that slightly longer periods of outage will be experienced.

The effects of congestion appear erratic. However, the average coverage, as well as the average and maximum outage times, generally increases with congestion up to approximately 15,000 cars as shown in Figure 1.9 to Figure 1.12. This is slightly counterintuitive; in previous cases when coverage has improved, outage time has decreased. This is due to the fact that increased congestion slows down the mobility of nodes in high traffic areas, such that they are more likely to be in contact for the required duration, hence improving coverage. At the same time, nodes that are out of coverage and caught in traffic are more likely to experience longer periods of outage.

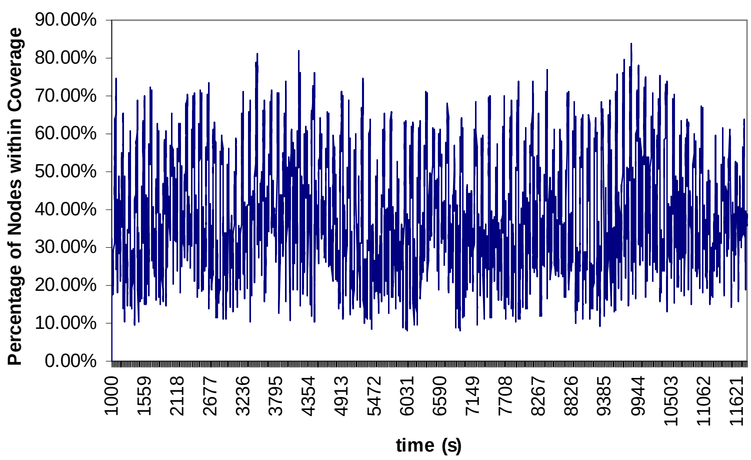


Figure 1.17: Typical Coverage with Congestion (30,000 cars)

Qualitative analysis suggests the results are attributable to non-uniform distribution of nodes and the mobility of nodes. Given that the destination selection mechanism for a movement sequence in our mobility model is an extension of that in the Random Waypoint Model, it is no surprise that the results exhibit Random Waypoint-like baseline behaviour. It has been shown that the asymptotic node density for the Random Waypoint Model is much higher at the centre of the simulation area but tends to zero at the borders, and the distribution is independent of node speeds [22, 23]. The placement of the taxi stand further contributes to the aggregation of the nodes. The non-uniform spatial distribution also applies to the non-taxi traffic, which mimics heavier traffic in the city centre. Meanwhile, mobility decreases under congested conditions, which generally reduces relative speeds. Therefore more nodes are likely to be close to the dispatch unit and hence reachable. Given the non-uniform distribution of congestion, however, nodes outside the highly congested city centre are able to move at higher speeds, and hence connectivity is poorer. On the other hand, the periodic variation is due to the density wave phenomenon observed for the Random Waypoint

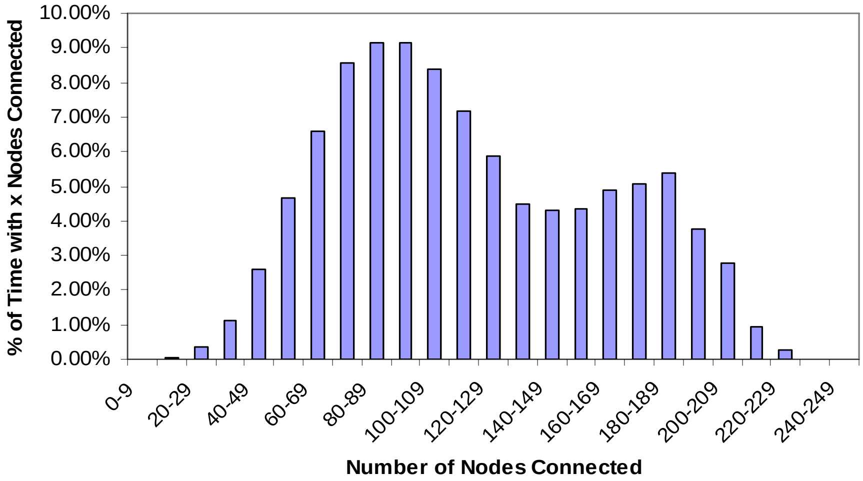


Figure 1.18: Typical Distribution of Coverage with Congestion (30,000 cars)

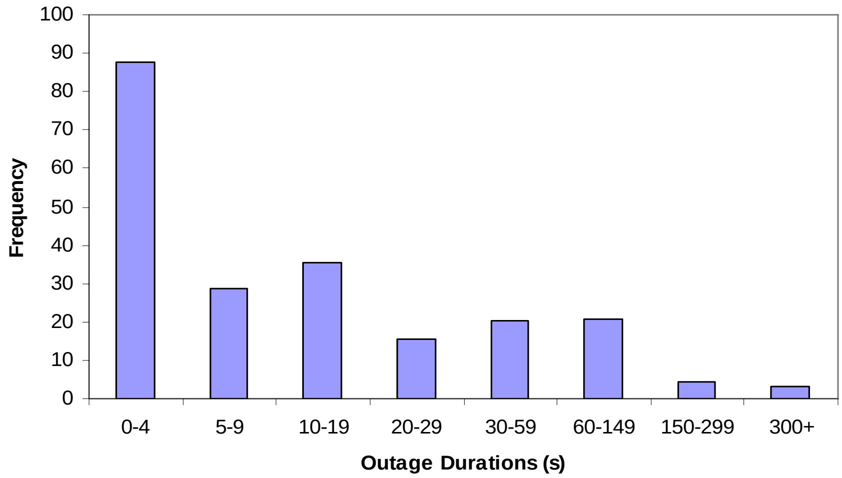


Figure 1.19: Typical Distribution of Outage Durations with Congestion (30,000 cars)

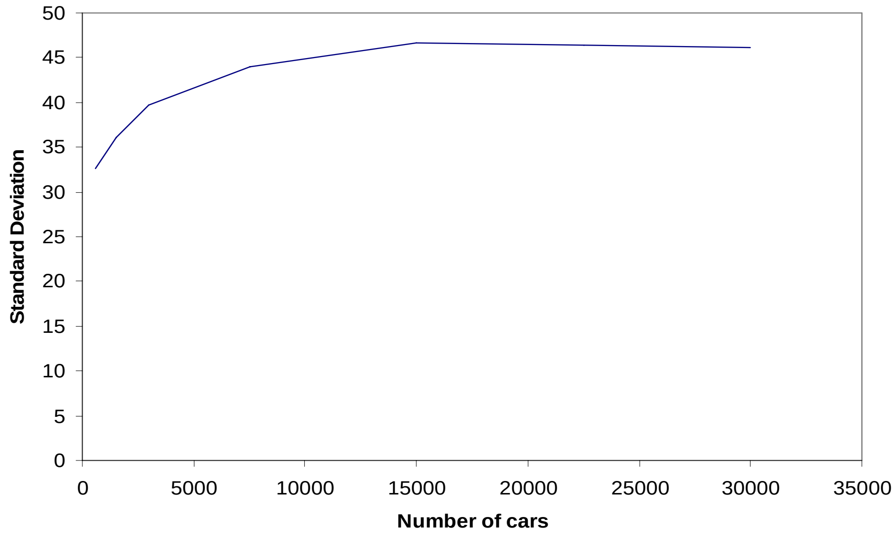


Figure 1.20: Standard Deviation of the coverage as Function of Congestion

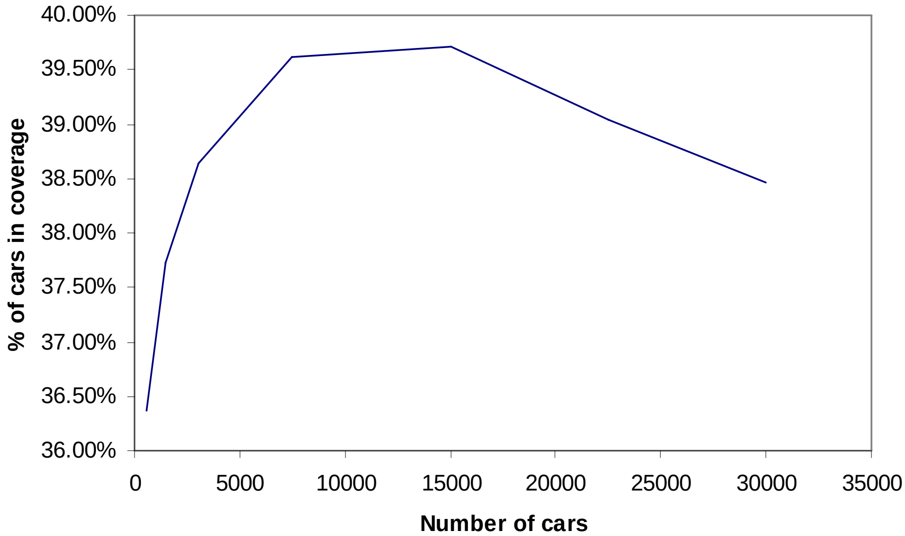


Figure 1.21: Coverage as Function of Congestion

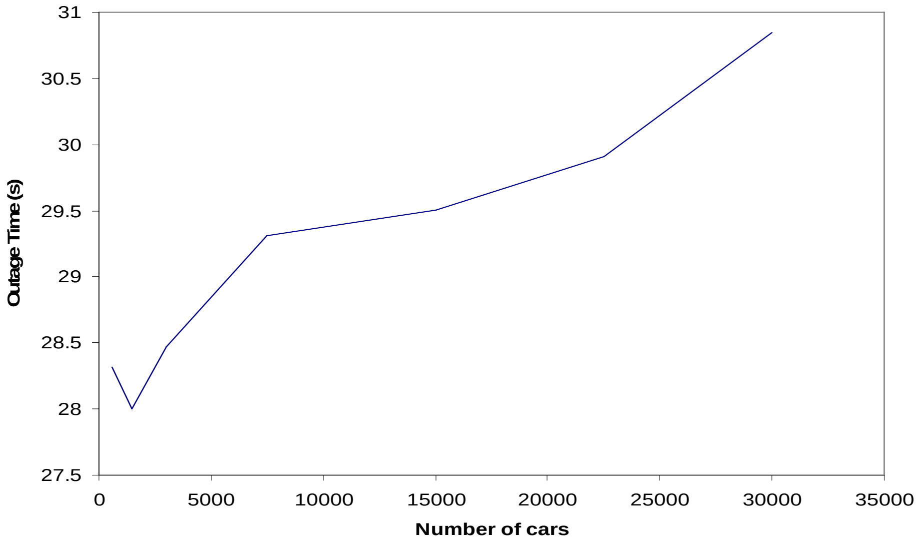


Figure 1.22: Average Outage Time

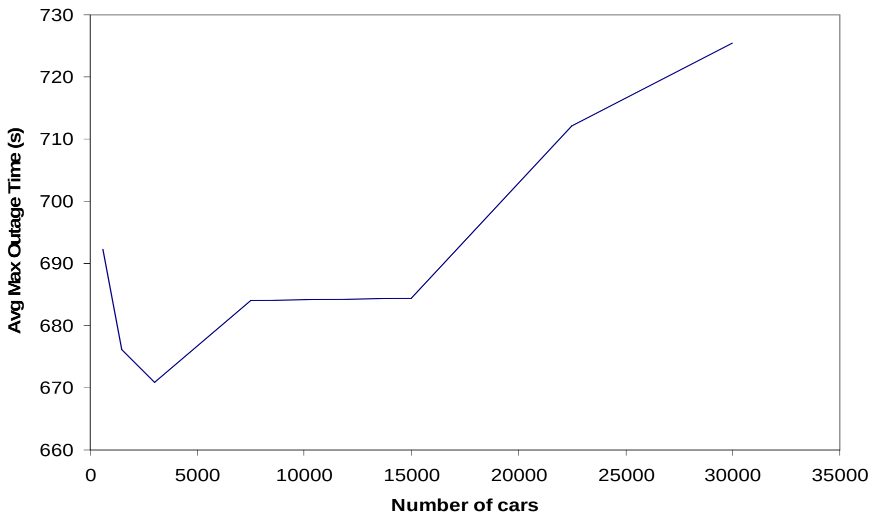


Figure 1.23: Average Maximum Outage Time

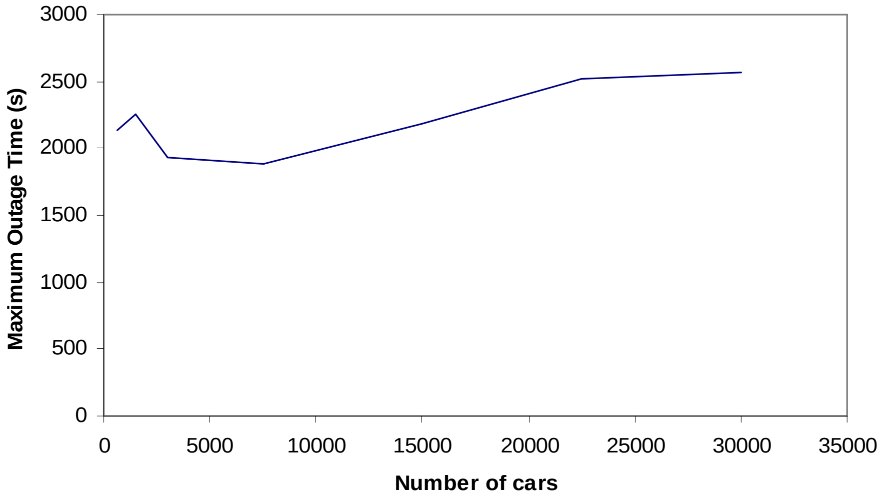


Figure 1.24: Maximum Outage Times

Model [9]. Similarly, our taxis are more likely to select paths through the central part of the simulation area, and therefore we can expect periodic dispersion of the nodes.

Above 15,000 cars, congestion is spread increasingly throughout the entire map such that all areas are now highly congested rather than just central, high traffic areas, and hence nodes are now more likely to be stuck in outlying areas, out of coverage. This subsequently results in a slight dip in average coverage and an increase in outage times as can be seen in the figures. Further investigations of the phenomena are under way to derive quantitative results.

These results suggest that with light to medium congestion the dispatch system will be able to operate on average as well or better than in zero congestion conditions. However, in conditions of very high congestion, coverage may suffer slightly and in any case congestion increases the risk of longer outages. In future work we intend to investigate the effect of events on mobility and coverage, (i.e., tidal effects, time-varying aspects) such as football matches where many taxis will be sent to a particular high traffic area.

1.3.3. Conclusion on Usage Scenarios

In this section, we have investigated the concept of a MANET based radio communications dispatch system. We looked at how the scenario might work and why it would be desirable as well as the risks and disadvantages of the system. By evaluating the system from both financial and technical perspectives, we are able to form a more complete picture of the systems feasibility.

While the concepts of vehicular or metropolitan area ad hoc networks are not new, our proposal addresses a specific application. And unlike conventional networks, our taxi network has a central management unit, even though communication is achieved in an ad hoc fashion. Through simulations, we highlight the application specific design considerations for ad hoc network applications. Given the particular information being transmitted, fit-for-purpose protocols and optimisations can be designed - e.g., use geographical routing towards or away from the central dispatch unit.

Such a system has the potential to bring value to all players involved, increasing the revenue per taxi, decreasing the cost and time to fulfil jobs and improving profitability per customer as well as customer satisfaction. The system would also be fast to deploy and scalable in size.

To evaluate the technical feasibility, we developed a realistic mobility and propagation model to establish the effects of node density, connection time and traffic congestion on various metrics of performance.

From the results, we deduce that whilst such a system would not be usable for real time communications, performance would be satisfactory for the purposes of the dispatch application. Since parameters and conditions in the simulation were meant to simulate harsh conditions, it is possible that in less difficult real world environments performance might be better than expected. However it is to be noted that performance also depends on factors such as routing protocols which are assumed to work as desired in this model. Whereas most research in the area tends to focus almost exclusively on these upper layers without considering the physical layer, we have through this work effectively identified an upper physical bound for connectivity in this network setting.

It is worth noting that we do not need to contact all taxis at all points. Therefore the performance results should not be interpreted in the conventional way. In particular, asynchronous routing would be most helpful in this case. As suggested by the outage times, the taxi network is highly intermittent, therefore the actual system would benefit significantly from a store-and-forward paradigm as in. Unreachable taxis are most likely unavailable too, due to taking passengers to outer areas of the city. Roaming taxis are likely to head for the central area anyway, as captured by the mobility model. Even when a free taxi is out of reach, it can still be flagged down on a street.

Given the unpredictable performance of the system, backup systems (such as mobile phones) may need to be in place for nodes which are stuck outside of coverage for an unusually long time.

It appears that QoS can be improved to suit individual specifications by increasing the number of nodes or decreasing the connection time or both. One option would be to introduce additional stationary nodes that double as taxi stands. This would also help to reduce the variability. Although a MANET based dispatch system appears to be both technically and financially feasible, there are a number of risks associated with it, in addition to the switching costs incurred by incumbent taxi companies. Given the incremental short term benefits, taxi companies may well choose not to adopt the new technology despite the potential long term cost savings. This could change depending on a variety of factors including technology and regulation. As spectrum pricing is currently under evaluation, it is possible that new regulations could make licensed bands more expensive, hence increasing the cost advantage for ad hoc networks. This could provide the necessary impetus to switch.

New operators would be able to adopt this system without worrying about switching costs, however the system will not be suitable for small companies with only a few cars until the system is standardised and competing companies can share the same relay system. A caveat is that standardising systems brings in additional issues to be resolved such as cooperation incentives.

The main alternatives to ad hoc systems are public or private radio networks. Research is also being performed to link conventional wireless LAN systems with seamless handoffs. The primary advantages that ad hoc networks have over these other forms of communication are low cost and ease of deployment. Should the alternatives advance sufficiently in performance or decrease substantially in price, they could render ad hoc systems inadequate in comparison for commercial uses.

On the technical side, security issues may raise concerns. Transmission must be secure to prevent eavesdropping of sensitive information. In the case of a standardised ad hoc network shared between different taxi companies, data encryption would avoid potential problems where companies attempt to poach each other's customers. Moreover, encryption is vital if credit card processing facilities are to be offered to customers. Situations may also arise where competitors try to perform denial of service attacks or otherwise disrupt communications to gain unfair competitive advantages.

Scalability issues bring up a further dimension of considerations. Both the software and the underlying radio technology must allow for significant growth in the number of nodes. With restricted bandwidth in a given area, there is a potential limit to the number of users that can be supported due to interference. This is unlikely to be an issue for taxi companies alone, but should the devices be sharing spectrum with other devices, problems could occur. One way to mitigate these problems would be to lower transmit powers or use adaptive/cognitive radios. Furthermore, opening up additional unlicensed bands could provide room for growth.

Regardless of whether the application is implemented in the proposed form, the main objective of this section is to highlight issues concerning the design of MANET based systems in city environments with high mobility nodes. The technical results discussed are applicable to any MANET application operating under similar conditions, as are many of the business concerns. An important conclusion from this work is that unless a critical mass of nodes can be deployed at the same time, a MANET based system could suffer from poor performance due to low node density. The expected number of nodes in operation needs to be considered when designing a system. Moreover, it is recommended that these systems be able to handle intermittent connectivity in light of the short but frequent outages experienced [27].

1.4. Mobility Models and Economics

Mobile ad hoc networking has been an active research area over the past decade, producing a multitude of protocols across the stack. Yet there remains much not well known of the network dynamics. This inspires us to carefully study the network characteristics to gain insight for protocol design. Due to the nature of the wireless medium and mobility, however, mobile ad hoc networks are highly dynamic and complex. Modelling the real world underpinning these networks is therefore difficult.

Models for ad hoc networks encompass views of the real world, including space configuration, radio frequency propagation and mobility, as well as network operations, such as communication patterns. We focus on the real world views. Despite much previous effort across the above aspects, the current state of art is still unsatisfactory. Most notably, criticisms centre on the lack of realism. While earlier models are credited with ease of understanding and implementation, they are often based on theoretical models rather than real world observations. More recent works have addressed some criticisms, and yet there remains much scope to improve the level of details, to more accurately and precisely reflect the underlying network.

Moreover, our work is encouraged by the importance of the models in simulations. Performance evaluations of ad hoc network protocols generally rely on simulations. These require either real world traces or appropriate models. The former is difficult to obtain, and often describe a specific setting, from which it is hard to derive general network characteristics. For the latter, the accuracy of the models is critical to the credibility of simulation results. It has been shown that the models used could significantly affect protocol performance. Therefore, an open question remains as to what effects those missing details might have. Our work considers these details.

The notion of space is embedded in most discussions on RF propagation or mobility, but we explicitly separate it out to emphasise its importance. Furthermore, criticisms and works have so far focused on individual propagation or mobility models, whereas the interplay between these models deserves much consideration. In this work, we propose a new integrated framework for capturing real world behaviour across different views and scenarios in the real world, MIRRORS: Mobility Integration of Radio Requirements in Real-world Simulations.

Although main discussions are in the context of simulations, we aim to model mobile ad hoc networks from a high level to deepen understanding of these networks. Therefore, instead of taking the common approach of focusing on evaluating existing protocols using these models, we use mobility as a focal point to study the topology characteristics of the networks synthesised. Our other contributions include: i) a new metric, neighbour occurrence count, to analyse the extent of intermittent connectivity in networks; ii) detailed distributions of the metrics studied; and iii) protocol design considerations highlighted from detailed modelling. An important observation is that these concerns could span several layers in the protocol stack. This could be of even greater interest and importance to cross-layer designs.

In the last part of this section we focus on a different aspect of mobility. Specifically, we look at mobility as an *opportunity for mobile devices to get in touch with each other*. This is a somewhat different way of looking at mobility, which is usually seen as a problem to deal with (e.g., mobility of nodes causes route breakage and service disruption), rather than as an opportunity. We introduce an *opportunistic networking* scenario, where contacts between mobile devices are exploited in an opportunistic way to send/forward information. Specifically, in Section 3 we characterise the contact time duration and the inter-contact times measured in a real environment. This is a first step to design networking solutions at the basis of real applications within such networking scenario.

1.4.1. Works related to the MIRRORS framework

The de facto standard

The set of models and parameters commonly used for performance analysis, especially of routing protocols, was originally proposed in [28]. The area is a flat unobstructed space of 1km x 1km. Radio propagation follows a two-ray ground reflection model, with a transmission range of 250 m, or 376 m for GloMoSim based simulations, over a perfect disc. Movement pattern is described by the Random Waypoint Model. The network size is often on the order of 50 nodes, and a typical simulation run lasts 900 s.

These settings certainly served as a good starting point for simulation based performance evaluations, though they do not match all scenarios. Unfortunately, they appear to have constrained later evaluations over years, which is probably unintended by the original authors.

RF propagation

Works on propagation models have largely involved link level measurements from test bed experiments, mainly as reported by Uppsala’s APE, Dartmouth’s experiments, MIT Roofnet and MSR’s experiments. Wide-ranging issues from connectivity to routing metrics reflecting link quality have been discussed, all highlighting the need for further real world experiments. However, in contrast to the common alleged MANET applications in battlefields, disaster relief and conferences, most test beds have been indoors and stationary. Roofnet is outdoors, but very high above the ground level. Dartmouth researchers recently conducted experiments of an outdoor mobile ad hoc network, but simulating Random Waypoint movements. It is also a pity that none of the works above has proposed empirical models matching the measurements.

Mobility models and metrics

As an integral part of the simulation model for mobile ad hoc networks, mobility models have been under extensive studies. Other than the Random Waypoint Model, several earlier random walk variants were described in [10], such as the Random Direction Model, where nodes randomly choose an initial direction for movements, as opposed to a random destination in Random Waypoint. Also described in the survey, Gauss-Markov Model correlates successive movements, although it is based on cell networks. Smooth Mobility Model includes detailed speed synthesis, especially for vehicles with high mobility, starting with a set of preferred speeds for each node. However, it is an improvement of Random Direction. The unconstraint environment is unlikely for the range of speeds, hence the types of nodes, studied. Mobility Vector Model focuses on natural speed dynamics, but does not address the overall trajectories.

More recent efforts have involved restricted movement spaces. The City Section Model in [10] considers a Manhattan grid like street network, but the formulation only dealt with predefined paths and traffic laws. The Freeway and Manhattan Models were proposed in [11]. In both models, nodes were constrained within lanes, and by the nodes ahead in the same lane. An additional consideration for the Manhattan Model was the turning behaviour at a junction, where a probabilistic choice is taken. This practice of applying the same probabilities at all junctions is debatable. Also in , a number of metrics were proposed to evaluate the mobility models, such as the degrees of spatial and temporal dependence. Their definitions ‘for nodes not too far apart in space/time’ were vague, however. The effects of obstacles were examined in [24], where a Voronoi space graph was used to generate paths from the obstacles. We will argue in the next section that obstacles and paths should not be coupled. In all three works, mobility models were shown to affect the relative ranking of the protocol performance.

In [29], the authors investigated the effects of destination selections on node densities and success rate of route discoveries. It should be noted that the nodes were restricted to students on campus, who do not roam continuously. Furthermore, the destinations have significant dimensions, when compared to waypoints, and the route discovery success is related to co-locations, one element of context.

The survey [19] further describes several group mobility models, which deal with movement dynamics from different angles. It was pointed out in [30] that group formation requires due consideration. The social aspects thus implied may be highly scenario dependent, but we demonstrate in the following on the work that our framework allows for these situations.

A number of works have considered mobility models from a statistical perspective, studying their stochastic properties, pointing out the problems with the original Random Waypoint Model and suggesting methodologies to ensure stationarity.

Where RF propagation is concerned, most models focus on path losses, and only has investigated signal obstruction. Even so, many other aspects, such as multi-path fading, have not been discussed but could make a difference, as shown in [31, 32].

In the case of mobility, constraints have been addressed in various models, but the fundamental movement mechanism is unclear, and the speed synthesis could be improved. Furthermore, none of the models incorporates all three streams of considerations.

It is worth noting that the geography is usually embedded in either RF propagation or mobility, as seen in the criticisms regarding flat free space. However, other aspects of the space geometry are left untouched but maybe significant.

Although mainly a simulation issue, the duration of a run is often limited to 900 s, except for [24]. The usual reason appears to be that the simulation will have converged to a steady state by then. However, it is questionable whether the full spectrum of behaviour is exhibited and whether stationarity is the only necessary concern. Later on we point out that minimum pauses could be longer than 900 s.

Most notably, all the improved models address one issue or two at a time, regarding RF propagation or mobility. Naturally, questions arise as to correlations between these issues, whether these enhanced models are comparable to one another, and if so, how?

1.4.2. The MIRRORS framework

Our first observation is that scenarios vary considerably, and therefore one single model is unlikely to suit all situations. In general, the space constrains RF propagation and mobility. Typically, the mobile nodes in question fall into a few categories: pedestrian, city vehicular and highway vehicular. Within each category, node movements exhibit similar patterns. Complex scenarios, on the other hand, can be decomposed into simple units. Same considerations apply across categories, in the form of same parameters albeit different values, e.g., pedestrians and cars alike may be under constraints from the geography. In some cases, however, some of the issues may simplify to varying degrees, e.g., accelerations are not of great concern to pedestrians. We therefore propose an integrated framework to address common concerns of space, RF propagation and mobility---MIRRORS: Mobility Integration of Radio Requirements Of Real-world Simulations.

The MIRRORS framework consists of a set of base cases, from which complex scenarios can be derived through composition. In each base case, nodes have homogeneous mobility capabilities and identical probabilistic distributions. Space is considered as a standalone component, while RF propagation and mobility models are other components of a base case. For each component, we specify parameters, the typical values of which are discussed in the context of representative base cases corresponding to the above categories. Effectively we group all parameters into different sets for space, RF propagation and mobility respectively.

Base case---space

The space model specifies3D geometries of the obstacles and paths, e.g., in the form of coordinates of obstacle vertices and control points for paths, and an appropriate projection onto a 2D area is often acceptable. The paths and obstacles are probably closely related, but not necessarily mutually exhaustive, considering, for example, lawns which neither favour movements nor block much signal transmission.

Base case---RF propagation

Given the common assumption of omni-directional antennas, propagation can usually be characterised by a micro cell environment. The main issues are path loss, fading and shadowing models. According to [17], appropriate empirical models can be used to simulate real world scenarios to a reasonable accuracy, and this was confirmed in [33]. For line-of-sight (LOS) paths, assumed on a flat surface, path loss can usually be approximated using the two-way ground reflection model and Ricean Fading is suitable. For non-line-of-sight (NLOS) paths, we consider that signals will be blocked by buildings, but diffracted around the vertical edges of buildings and potentially over rooftops. Fast fading is more like to conform to Rayleigh statistics. The signal power from the LOS path with respect to the power from NLOS paths can be controlled by the Ricean K Factor. Shadowing is useful to model the time varying signal strength due to, e.g., leaves.

In most cases, it requires applying well known empirical models, and our emphasis is on considering all applicable components. RF propagation is highly complex, and therefore it is advisable to note both the merits and limits of simulation approaches.

Base case---mobility

For ad hoc networks, we model human movement behaviour, unlike for sensor networks for environment monitoring. Mobility is continuous, in contrast to mobile IP, cell network or

WLAN situations, where mobility is discrete. Furthermore, we are concerned with microscopic behaviour, since movement dynamics is down to individual nodes. Therefore, we focus on entity mobility models. Compared with space and propagation models, mobility models are less well defined, and hence this is our focal point.

A number of general observations can be made. Nodes normally follow targets, and movements are under constraints. People prefer to save time, by travelling at a speed that is close to the possible maximum while ensuring a comfortable state for themselves.

The building blocks for a movement trace include movements and journeys. We define each movement as a period of motion at constant velocity, and each journey as a sequence of movements from the last destination up to the current target. Pauses might be possible between journeys. Concatenating journeys yields the overall trajectory, while instantaneous velocity is the rate of change along the trajectory, as in standard physical definitions.

It is the trajectory that dictates a node’s trace asymptotically, while the detailed speed variation determines the precise point along the trace at a particular instant. To that end, trajectories and speed variations can be derived separately.

Each node is confined in a movement space, which may be an open campus for pedestrians, city streets or freeways for cars. In practice all nodes in the network are likely to be subject to the same movement space, and hence this could be a parameter across the whole simulation.

A node has a list of preferred destinations and other destinations, which together specify the destination distribution for a journey. In general, this distribution is dependent on the node’s current location and time, and can be described by a Markov Chain.

Other parameters include characteristic speed(s), preferred/steady state speeds, associated speed drifts, speed limits, acceleration limits, altogether describing the speed dynamics. Compared with the Smooth Mobility Model, we assign only one preferred speed to each node, and leave other speeds synthesised through constraints.

Not all parameters are of concern in each scenario, e.g., in a pedestrian network it suffices to consider speeds without regard to accelerations. Furthermore, a pause time distribution is associated with each location and time. It can be short to represent roaming, or long to reflect the more common behaviour of travelling to a place and staying for a long time.

To start a new journey, a node selects a new destination according to the destination distribution at its current position and time. A movement algorithm then describes how the node travels to the destination. Along the trace, the velocity of the node is adjusted according to spatial, temporal and physical constraints and speed drifts, as well as any issue particular to the scenario. Node positions are updated accordingly. On reaching the destination, the node possibly pauses. Then the whole process is repeated.

Dependencies between models

From the viewpoint of mobility, the characteristic speed identifies the category of movement patterns, and hence the typical space. A key observation is that, for a particular scenario, the space is the underlying substrate for the network. This is in contrast to the wire line networking paradigm where space can often be abstracted away in the presence of wires, and it may be appropriate to abstract them further by capturing link metrics. Generally, the paths constrain line-of-sight RF propagation and mobility, whereas obstacles affect signal obstruction and diffraction. The introduction of an obstacle would therefore affect both the movement freedom and the signal propagation, as observed in . Furthermore, approximations in RF propagation calculations may depend on the mobility scenario, especially when concerning fading and shadowing situations.

Therefore, a scenario involves not only the mobility and the movement space, but implicitly the propagation. A consistent approach is necessary in modelling, which is also suggested by our framework.

A number of statistical issues are of concern to simulations. For example, statistical artefacts due to small area sizes have not received sufficient attention. This is one result of the common approach of modelling a closed system. Considering large systems, most are in fact self-contained, so a closed system can be a reasonable representation. To impose ‘closedness’ on a small area necessitates border rules, and previous studies have shown that their impacts are significant. On the other hand, a small area limits the extent of dispersion, and therefore may not present the full spectrum of connectivity paradigms, from persistent, intermittent to transient, corresponding to a node in contact with the same neighbour continuously over a long time, or encountering the neighbour recurrently or occasionally. Also shown in various studies, the spatial distribution of nodes is normally non-uniform.

In order to reflect these, the configuration of the simulation area needs to comply with realistic vision. To model a metropolitan area, for example, the scale should be at least several kilometres on each side. In certain cases, we should allow nodes to depart and arrive temporarily, i.e., allowing the total number to vary but within a tolerance. Depending on the area size, the simulation duration should be sufficiently long to ensure convergence both speed- and destination-wise. The border behaviour, initialisation and position update procedures will also vary with the scenario. In summary, the movement scenario should determine the simulation scale.

As mentioned earlier, mobile nodes are normally pedestrian, city vehicular or highway vehicular. Together with the case of stationary nodes, they form categories of ad hoc networks. We now discuss representative patterns.

Stationary network

The characteristic speed is 0, and this is usually a network of indoor devices. It may be generalised to cover fixed base stations on a wide area, although a standalone network of those is unlikely. RF propagation is complicated and under extensive studies, but the absence of mobility reduces the network dynamicity considerably. A stationary network may be more concerned with mesh connectivity and capacity, but within this framework it still serves as a base case.

Pedestrian network

The characteristic speed is the maximum speed, 2 m/s. Typical pedestrian networks are in campus environments or metropolitan areas. In either case, the movement space is 2D with respect to the paths, and obstacles in the paths are ‘points’, mainly building and trees. This means that nodes are generally not under constraints from peers, unless the entire area is saturated. RF propagation involves path losses along the paths, obstruction by obstacles, diffractions around vertical edges of obstacles, fading and shadowing concerns.

In terms of mobility, the key is destination distribution, which is non-uniform across time and space. Different groups of people would form different cases. Consider students on campus, they tend to aim for residence halls, departments, libraries, canteens and so on, and the preferences vary according to time of the day. Professors would have different preferences, but similar considerations apply. This generalises [29], which is based on students’ movements. Most likely the person would travel at a preferred speed he or she is comfortable with, and there is little variation beyond that. The pause time is associated with the purpose of the journey, such as meals in canteens and lectures in departments, hence the location, and is likely to be long, e.g., at least 20 minutes. Therefore these nodes do not roam. This can be contrasted with pedestrians in a shopping mall, where they tend to pause only briefly and roam for a considerable amount of time. Other issues or constraints, such as accelerations, are negligible.

City vehicular network

In this case the characteristic speed is the city-wide speed limit. The movement space consists of lanes in streets, and therefore movements are linear with respect to the paths. For RF propagation concerns, path losses are similar to those in the pedestrian cases, but fading/shadowing effects can often be ignored due to the high mobility.

Representative mobile nodes in these scenarios are buses, taxis and other traffic. Buses roam, but have fixed trajectories. Their destinations are always the next stops along the routes, where they pause very briefly or occasionally not at all, so the movements are mostly deterministic. Taxis also roam, but the traces are more random and pervasive. They tend to favour places like bus/train stations or other busy areas in town, but may also travel to other destinations to collect or drop off a passenger. The pause times tend to be short and above a non-zero minimum, but special considerations might apply at taxi stands, e.g., if the taxi joins a queue and waits for its turn to take the next passenger. Other vehicles often have transient journeys, potentially pausing for a significant length of time after a journey, and make fewer discrete journeys over a long period of time, on the order of days.

Speed-wise, in any of the three cases the driver usually prefers a speed close to the speed limit, and the actual speed drifts slightly around the preferred value under free flow conditions. The constraints include spatial---safety distance to vehicles ahead in the same lane and traffic lights, temporal---velocity correlations between successive movements in a journey, and physical---acceleration bounds in general and speed bounds for turning action and so on. Given the high mobility setting, accelerations are significant. Road policies are examples of additional concerns for this scenario, e.g., considering one-way streets, which could affect the initial movement in a new journey and turning actions.

Highway vehicular network

This setting is characterised by a higher speed limit than in built-up areas. The space is usually open and uncluttered, which simplifies both radio and mobility concerns. Movements are essentially continuous along the lanes, and the same speed variations as above apply. The main subtlety is in switching lanes, for which many driver behaviour models have been proposed for transport studies.

Heterogeneous scenarios

All the representative patterns above depict very homogeneous networks, where nodes have the same mobility capabilities and identical probabilistic distributions. They can serve as the building blocks for heterogeneous situations.

Heterogeneous space

An interesting example of space composition is to combine indoor and outdoor environments over wide area, e.g., consider the vicinity of a campus building as well as the inside. More generally, non-uniform space layout could be derived from basic area units. Essentially, ‘constituent states’, positions or signal strengths, can be calculated within the respective space components, and care is needed for dealing with behaviour along component boundaries.

Heterogeneous radio models

Since the basic propagation models already embrace a number of issues, composition is not as distinct. Some are implied in space composition examples, e.g., indoor/outdoor environments would imply different path loss exponents, and therefore they should be applied as appropriate. It may however be possible to accommodate single-radio nodes with variable transmission ranges or multi-radio cases, by treating them as compositions of single radios with fixed ranges.

Heterogeneous mobility

If a group of nodes follow the same destination selection pattern, then group mobility could be observed. This further suggests that the choice of destinations could embody social aspects in the network. The notion of co-location in [30] is reflected in neighbouring relations, i.e., nodes within direct connectivity are likely to be close in location. Given the vast number of possibilities in resulting in group mobility, it can be more scenario dependent than entity cases.

Heterogeneous networks can be derived if the requirements of homogeneous node capabilities and identical probabilistic patterns are relaxed. One example is non-identical patterns within the same types of nodes, e.g., considering arts students and science students, whose destination distributions could be the same, except they involve different departments. Within the same category of mobility capability, we observe city traffic networks with buses, taxis and normal traffic, which have different behaviour. Crossing capability boundaries, we could derive the cases of mostly stationary networks with occasional mobility, as in many test bed experiments, and metropolitan area networks with both pedestrians and vehicles. Higher order compositions of composed models could generate even more complex patterns.

Despite the increasing degree of heterogeneity, the mobility states of nodes, such as destinations and current positions, can be obtained from respective mobility units, with additional adjustments as a result of interactions between different nodes. Where different movement space is involved, RF propagation calculations need potential corrections accordingly. The viability of such heterogeneous networks for the purposes of network services is beyond the scope of this discussion, however.

Comparative mobility model analysis using the City Taxi Scenario

A number of mobility models have been proposed for MANET simulations, such as the Random Waypoint [8], Random Direction [9] and other Random Walk variants [10], amongst which the Random Waypoint is the most widely used. Despite its popularity, the Random Waypoint Model lacks both realism and desirable statistical properties [25]. Although there have been recent efforts to design more realistic models [24], the level of detail is insufficient for our feasibility evaluation. For this reason we did not use packages such as ns-2/GloMoSim, as their models were too simple for our purposes - delete. However, our simulation is compatible with them and our mobility traces can be fed into ns-2 using the setdest tool.

Some simulations have used real life bus traces to model general vehicular mobility. However, bus movements are fairly regular and periodic, and only cover certain parts of the city, whereas taxis and other vehicles tend to have somewhat random destinations and pervasive trajectories. Furthermore, as with most studies, the physical layer was not considered, and it was assumed that signals had a 1.5 km radius range despite the city block environment, which may be overly optimistic. We model the city as a Manhattan style grid, as in [11], with a uniform block size across the simulation area. All streets are assumed to be two-way, with one lane in each direction. Taxi movements are constrained by these lanes, which are modelled using one-metre segments.

We assume that, under free flow conditions, each taxi's behaviour is homogeneous with respect to itself over time. A taxi is characterised by a preferred speed, a maximum acceleration and deceleration [12], a speed variation associated with the preferred speed at steady state, and a list of preferred destinations, i.e. the taxi stands. The taxis are randomly assigned one of three preferred speeds. All other parameters are set to be the same across all taxis, as they are from the same company.

At any instant, a taxi is either i) carrying a passenger to a destination, assumed to be within the city, ii) heading for the taxi stand, or iii) roaming around until flagged by a passenger. The third case can be viewed as a taxi travelling to a particular location to collect a passenger. We combine into the roaming case the situation where a taxi collects a passenger according to the dispatch unit. Therefore, taxi journeys comprise of a sequence of movements of constant velocities, starting from the current position and ending at the destination. In the first two cases, the taxi pauses on reaching its destination; the roaming situation can comprise of several journey sequences with smooth transitions, depending on whether the taxi is to pause. The probability of a taxi picking up a passenger is estimated from a real taxi's daily empty cruising time.

Given parts of the town is more popular/dense, perhaps due to shopping area, etc. though they may be pedestrian areas. From a taxi's viewpoint, we know from experience that taxis tend to roam to places with high aggregation flows - train stations, coach stations, airports, etc. given their popularities are pretty similar (this varies from city to city), we consider the probabilities of any of them being the destination are the same across places and taxis (more explanation however - e.g., consider the train stations during rush hours, maybe more people arrive, but commuters aren't as likely to take a taxi as a passenger who arrives at other times of the day, who probably comes with more luggage and would need a taxi.

To start a sequence of movements, a taxi selects a destination according to the following distribution: probability of 0.5 uniformly shared between all taxi stand(s), and otherwise uniform within the remaining 0.5 probability for any other position in the simulation area. Therefore if there is only one stand, there is a 50% chance that a taxi will head towards it, otherwise there is an equal chance of it moving towards any other position on the map. We assume that these taxi stands are located in popular areas that see particularly high traffic such as train/coach stations to account for hot spots around the city. Adjustments to the selection procedure ensure that the destination is not the current position. The movement algorithm in principle computes a shortest path towards the destination. Since such a path is not unique in the grid layout, the probability of taking each direction at a junction is determined according to the distance components to the destination. At the borders the algorithm specifies 'bounce-back' behaviour. On reaching the destination, the taxi either pauses according to an exponentially distributed pause time plus a minimum 30 seconds, or continues roaming into the next movement sequence. An adjustment is made when the destination is a taxi stand, where the taxi joins a queue and waits for its turn to take the next passenger. The selection of destinations as opposed to initial directions [9, 12] better represents the true nature of real world movement towards a target.

Successive movements within a sequence are correlated by the taxi's velocities [12, 13]. The direction will be along the same lane until a possible turn at the nearest junction. The speed variation is bound by the acceleration or the deceleration [14], and if at steady state, it is specified by the steady state variation which simulates speed drifts. This also ensures that a taxi speeds up and slows down gradually, respecting physical laws [12, 15], at appropriate times. Further constraints on the taxi movements are imposed by the speed limit of the lane [12] (set to be 15m/s), previous taxis in the same lane [11] and traffic lights. We assume traffic lights are installed at each junction, with the green and red lights on for 60 seconds alternately. All lights along an entire street are consistent. For simplicity, traffic lights do not affect taxis about to turn into another street.

At the start of a simulation run, the taxis are placed according to their destination distribution, and their speeds set to the preferred speeds, with adjustments for those at junctions. We believe this approximates a steady state setup, and will allow the simulation to converge quickly [16]. Taxi positions are updated at one-second time steps to approximate continuous motion. Given the potential inter-dependency between movements of taxis, e.g., two taxis turn into the same lane from two directions, one constraining the movements of the other, they are updated in a randomised order at each second.

The simulation models a MANET based on the IEEE 802.11b standard, since they are well understood, readily available and adaptable to the application under consideration. 802.11b operates in the 2.4 GHz ISM band, and we assume a typical receiver sensitivity of -80 dBm and a transmit power of 100mW (20dBm). Since our system will operate using taxis' electrical systems as opposed to mobile nodes powered by batteries, power consumption is not a big issue.

The results of the simulation will depend heavily on the environment that the nodes operate within. In this case, it is our intention to simulate harsh conditions in order to test the feasibility of the scenario. As such, we have used a Manhattan grid model where transmission is limited by blocks of buildings placed at regular intervals.

As our omni-directional antennas will be mounted atop the roofs of taxis at a height of around 1.5m, propagation will be characterised most accurately by a micro cell model. The dominant propagation mechanisms in such an environment are due to interactions between the direct path and paths reflected from buildings and the ground. Given the low heights of our antennas, we ignore diffraction over rooftops.

We also ignore interference from other nodes in the model as the density of nodes is very low. Should node density increase, the distance from node to node will decrease and the transmit power can simply be lowered to shorten the propagation range, effectively maintaining or even increasing the bandwidth available. Coupled with the low heights of the antennas this leads to very small reuse distances, thereby minimising problems from interference.

Our propagation model is a dual slope empirical model appropriate for path losses in micro cells [17]. Essentially this model assumes a 2nd order loss (-20dB/dec) out to a 'break' distance and a 4th order loss (-40dB/dec) thereafter. For antenna separations greater than the 'break' distance, the plane earth propagation model yields,

$$\frac{P_r}{P_T} = \left(\frac{h_1 h_2}{d^2} \right)^2$$

where Pr is the received power (W), PT is the transmit power (W), h1 and h2 are the node antenna heights (m) and d is the antenna separation (m). The breakpoint distance is approximately given by,

$$d_b = \frac{4\pi h_1 h_2}{\lambda}$$

and is the point after which the ground reflection destructively interferes with the direct ray and reduces the field strength. For the simulation, the 'break' distance is set at 100m. Both nodes are assumed to be in line of sight.

Thus, before the breakpoint we have,

$$Gain(dB) = -20\log_{10} d - 32.9563 \tag{1}$$

and after the breakpoint,

$$Gain(dB) = 20\log_{10} h_1 + 20\log_{10} h_2 - 40\log_{10} d \tag{2}$$

To model the diffraction loss at the corner of buildings, we estimate that every time a corner is taken, a loss of 20dB is incurred in addition to the losses in equations 1 and 2. We also assume digital transmission via a narrowband channel. In other words, multipath propagation is assumed to be sufficiently small that the reception and detection of transmitted data will not be degraded by its presence. This is reasonable because the small coverage areas of micro cells and line of sight paths lead directly to a reduction in multipath delay spread; fast fading data is most likely to conform to Rician rather than Rayleigh statistics [18] also can be ignored at high mobility.

Using the conditions specified, the maximum unobstructed transmit distance is 474m along a straight line and 150m if one corner is taken.

Operations in the system are based on simple message exchanges. Taxis regularly send status updates to dispatch, and dispatch sends job requests and other information to taxis. Messages are always sent to the nearest reachable hop in order to minimise transmit power and interference. Given that we know the location of dispatch and any taxi stands, as well as the taxis themselves, the system could use a location aware routing protocol such as location-aided routing (LAR), geographic distance routing, grid, zone-based two level routing or a customised derivative. However, routing and MAC layer issues are beyond the scope of this paper, which is concerned mainly with the physical layer feasibility of such a mobile ad hoc network. It is assumed that upper layers are functioning appropriately; as long as a signal of sufficient strength is received we assume that communication of acceptable QoS can be achieved. By taking a purely physical viewpoint, we are able to determine whether connectivity is physically achievable, independent of the performance of upper layers, which tend to vary considerably.

Being an ad hoc network, it is likely that taxis will wander in and out of coverage, the frequency and duration of which depend on the number of relay points, transmission range of each device, movement of vehicles, average distance between cars, etc.

However, as long as the period of outage is not too great, this may not necessarily impact performance significantly. Although real time communications require constant connectivity, a dispatch system does not need to transfer job data instantaneously and can operate on an asynchronous messaging basis. Since the amount of data required is very small (names, pickup and destination addresses, special instructions etc), data can be transmitted quickly whenever nodes are within coverage. Therefore as long as outage periods are short, the fact that connections are not constant should be of minimal concern to users and performance results should be interpreted with this in mind.

Even when the amount of data required is large, from similar types of studies, it is possible that in conditions of intermittent connectivity, a buffer or drive-thru proxy [19, 20] (based at dispatch, taxi stands and the individual nodes themselves) could be used to accumulate data and requests/responses from peers within the network and forward the information at the next opportunity without significantly sacrificing performance.

The simulation models used in mobile ad hoc network research have been criticised for lack of realism. While credited with ease of understanding and implementation, they are often based on theoretical models, rather than real world observations. Criticisms have centred on radio propagation or mobility models.

In this work, we take an integrated approach to modelling the real world that underlies a mobile ad hoc network. While pointing out the correlations between the space, radio propagation and mobility models, we use mobility as a focal point to propose a new framework, MIRRORS, that captures real world behaviour. We give the formulation of a specific model within the framework and present simulation results that reflect topology properties of the networks synthesised. Compared with the existing models studied, our model presents a wider spectrum of variation in the metrics examined, due to the model encapsulating more detailed dynamics. While the common approach is to focus on performance evaluation of existing protocols using these models, we discuss protocol design opportunities across layers in view of the simulation results.

Concluding remarks on the MIRRORS framework

We have proposed a new integrated framework, MIRRORS, that models real world ad hoc networks. This integration can be seen as across space, RF propagation and mobility, across nodes with different mobility capabilities and with consideration to both protocol design and evaluation. By presenting comparative results, we highlight fine variations within the network dynamics captured by the details in our model.

Parameters have been identified for models within our framework, and appropriate values could be extracted from a modest amount of real world traces to obtain the precise models.

Despite the microscopic focus, macroscopic behaviour of the networks could be derived from the models within the framework. This could facilitate fit-for-purpose protocol design or prescriptive optimisations of existing protocols. We call for a synergy between realistic modelling and protocol design, noting that a real world scenario impacts all layers in different ways. The set of representative cases in the framework could also serve as a test-suite to evaluate the performance of a general protocol.

1.5. References

[1] Marc Bechler, Walter J. Franz and Lars Wolf, February 2003, “Mobile Internet Access in Fleetnet”, Verteilten Systemen KiVS 2003, Leipzig, Germany

[2] E Huang, S Östring, J Crowcroft & I Wassell, “Developing Business Models for MobileMAN”, Technical Report CUED/FINFENG/TR.475, Cambridge University Engineering Library, 2004 (omit for now)

[3] John R. Stone, Gorman Gilbert, and Anna Nalevanko, “Assessment of Computer Dispatch Technology in the Paratransit Industry”, U.S. Department of Transportation (monetary figures have been converted to current day values using http://eh.net/hmit/ppowerusd/)

[4] Y. Cheng, “Singapore’s Smart Taxis”, Asian Technology Information Program (ATIP), March 1996

[5] Interview with with director of a local taxi company

[6] Dolphin Telecommunications homepage, http://www.dolphin-telecom.co.uk

[7] Cordic Communications website, www.cordic.com

[8] J. Broch, D. A. Maltz, D. B. Johnson, Y. C. Hu and Jorjeta jetcheva, “A performance comparison of multi-hop wireless ad hoc network routing protocols”, in ACM/IEEE MobiCom, 1998

[9] Elizabeth M. Royer, P. Michael Melliar-Smith, and Louise E. Moser, “An analysis of the optimum node density for ad hoc mobile networks”, IEEE ICC, June 2001

[10] Tracy Camp, Jeff Boleng and Vanessa Davies, “A Survey of Mobility Models for Ad Hoc Network Research”, in Wireless Communication & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications, vol.2, no.5, 2002

[11] Fan Bai, Narayanan Sadagopan and Ahmed Helmy, “IMPORTANT: A framework to systematically analyze the Impact of Mobility on Performance of RouTing protocols for Adhoc NeTworks.” in IEEE INFOCOM, March 2003

[12] Christian Bettstetter, “Smooth is Better than Sharp: A Random Mobility Model for Simulation of Wireless Networks”, MSWiM’01, July 2001

[13] Ben Liang and Zygmunt J. Haas, “Predictive Distance-Based Mobility Management for PCS Networks”, in IEEE INFOCOM, March 1999

[14] Christian Schindelhauer, Tamás Lukovszki, Stefan Rührup and Klaus Volbert, “Worst Case Mobility in Ad Hoc Networks”, in SPAA’03, June 2003

[15] Xiaoyan Hong, Taek Jin Kwon, Mario Gerla, Daniel Gu and Guangyu Pei, “A mobility framework for ad hoc wireless networks”, in ACM International Conference on Mobile Data Management (MDM), Jan 2001

[16] Jungkeun Yoon, Mingyan Liu and Brian Noble, “Sound Mobility Models”, in ACM/IEEE MobiCom, Sept 2003

[17] Simon R. Saunders, WILEY 1999, “Antennas and propagation for wireless communication systems”

[18] E Green, “Radio link design for microcellular systems”, British Telecom Technology Journal Vol 8 No 1 Jan 1990

[19] Mattias Esbjornsson, Oskar Juhlin and Mattias Ostergren, Dec 2002, “The Hocman Prototype-fast motor bikers and ad hoc networking”, International Conference on Mobile and Ubiquitous Multimedia

[20] Jorg Ott and Dirk Kutscher, May 2004, “The “Drive-thru” architecture: WLAN-based internet access on the road”, IEEE Semiannual Vehicular Technology Conference

[21] Ingo Gruber, Oliver Knauf, Hui Li “Performance of Ad Hoc Routing Protocols in Urban Environments”, Proceedings of European Wireless 2004 (EW’2004), Feb 2004

[22] Douglas M. Blough, Giovanni Resta and Paolo Santi, “A Statistical Analysis of the Long-Run Node Spatial Distribution in Mobile Ad Hoc Networks”, in MSWiM’02, Sept 2002

[23] Christian Bettstetter, Hannes Hartenstein and Xavier Pérez-Costa, “Stochastic Properties of the Random Waypoint Mobility Model”, in ACM/Kluwer Wireless Networks, Special Issue on Modeling & Analysis of Mobile Networks, Vol. 10, No. 5, Sept 2004

[24] Amit Jardosh, Elizabeth M. Belding-Royer, Kevin C. Almeroth, Subhash Suri, “Towards Realistic Mobility Models For Mobile Ad hoc Networks”, in ACM/IEEE MobiCom, Sept 2003

[25] Jungkeun Yoon, Mingyan Liu, Brian Noble, “Random Waypoint Considered Harmful”, in IEEE Infocom, April 2003

[26] David D. Clark, John Wroclawski, Karen R. Sollins, Robert Braden, “Tussle in Cyberspace: Defining Tomorrow’s Internet”, in ACM SIGCOMM, August 2002

[27] Kevin Fall, “A Delay-Tolerant Network Architecture for Challenged Internets”, in ACM SIGCOMM, August 2003

[28] J. Jetcheva, Y.C. Hu, S. PalChaudhuri, A. Saha, D. Johnson, “Design and Evaluation of a Metropolitan Area Multitier Wireless Ad Hoc Network Architecture”, in Mobile Computing Systems and Applications, 2003

[29] K. Merchant, W. Hsu, H. Shu, C. Hsu and A. Helmy. Weighted way mobility model an its impact on ad hoc networks (poster). In Proceedings of ACM/IEEE MobilCom, 2004.

[30] M. Musolesi, S. Hailes, and C. Mascolo. An ad hoc mobility model founded on social network theory. In Proceedings of ACM MSWiM 2004.

[31] D. Aguayo, J. Bicket, S. Biswas, G. Judd and R. Morris. Link-level measurements from an 802.11b mesh network. In Proceedings of SIGCOMM 2004.

[32] M. Takai, J. Martin and R. Bagrodia. Effects of wireless physical layer modeling in mobile ad hoc networks. Proceedings of ACM MobiHoc, 2001.

[33] R. Gray, D. Kotz, C. Newport, N. Dubrovski, A. Fiske, J. Liu, C. Masone, S. McGrath, and Y. Yuan. Outdoor experimental comparison of four ad hoc routing algorithms. In Proceedings of ACM MSWiM 2004.

.

2. MESH NETWORKS: COMMODITY MULTI-HOP AD HOC NETWORKS

2.1. Introduction

In spite of the massive efforts in researching and developing mobile ad hoc networks in the last decade, this type of networks has not yet witnessed a mass-market deployment. This opened a debate in the scientific community on why, after almost a decade of research into ad hoc networking, MANET technology has not yet affected our way of using wireless networks. A common answer is emerging[1]: most of the ongoing research on mobile ad hoc networks is either driven by Department of Defense (DoD) requirements (large scale military applications with thousands of ad hoc nodes) or by specialized civilian applications (disaster recovery, planetary exploration, etc). DoD generated a research agenda and requirements that are far from real users requirements. Indeed, military and specialized civilian applications require lack of infrastructure, and instant deployment. They are tailored to very specialized missions, and their cost is typically not a main issue. On the other hand, from the users' standpoint, scenarios consisting of a limited number of people wanting to form an ad hoc network for sharing some information, or an access to the Internet, are much more interesting. In this case, users are looking for multipurpose networking platforms in which the cost is an issue, and Internet access is a must. To turn mobile ad hoc networks in a commodity some changes to the original MANET definition seem required. By relaxing one of the main constraints of mobile ad hoc networks "the network is made of users devices only and no infrastructure exists" we move to a more pragmatic networking scenario in which multi-hop ad hoc networks are not isolated, self-configured networks, but rather they will emerge as a flexible and "low cost" extension of wired infrastructure networks and will coexist with them. Indeed, a new class of networks is emerging from this view: the *mesh networks* [2]. Mesh networks are built upon a mix of fixed and mobile nodes interconnected via wireless links to form a multi-hop ad hoc network. As in MANETs, the users' devices are an active part of the mesh. They dynamically join the network, acting as both users' terminals and routers for other devices, consequently further extending the network coverage. Mesh networks thus inherit many results from the MANET research but have civilian applications as the main target. Furthermore, while the MANET development approach was mainly simulation-based, from the beginning, mesh networks have been associated with real test-beds. By designing/implementing "good enough" solutions it has been possible to verify the suitability of this technology for civilian applications and to stimulate the users' interest to adopt it. Even though mesh networks are a quite recent technology they have already shown big potentialities in the wireless market. Indeed, we can subdivide the mesh networks in two main classes: off-the-shelf and proprietary solutions. An example of the first class are the so-called *community networks* built (mainly) upon 802.11 technology and aimed at providing Internet access to a community of users that can share the same Internet access link [3]. Several examples of this are: Seattle Wireless, Champaign-Urbana Community Wireless Network (CUWiN), the San Francisco BAWUG, the Roofnet system at MIT (MIT Roofnet). On the other hand, several companies are now selling interesting solutions that exploit the mesh network potentialities for indoor and/or outdoor applications (e.g., MeshNetworks, Tropos Networks, Radiant Networks, Firetide, BelAir Networks, Strix Systems, etc.)[2]. For example, indoor mesh networks can be set up by wireless interconnected access points that, by exploiting routing algorithms developed for MANETs, can create extended WLANs without a wired infrastructure. Outside buildings, meshes networks can be used to provide wireless

[1] See, for example, Mario Gerla in the IEEE MASS 2004 panel, and Victor Bahl in the opening talk at the Mesh Networking Summit 2004

[2] For an exhaustive list of web links related to ad hoc and wireless mesh networks, the reader could refer to http://www.antd.nist.gov/wctg/manet/adhoclinks.

access across wide geographic areas by minimizing the number of *wired* ingress/egress points towards the Internet. Outdoor networks might be used, for example, by municipalities to extend their wired networks wirelessly.

This promising networking technology recently received a further boost by IEEE 802 that created the Task group 802.11s aimed at defining a MAC and PHY layer for meshed networks to improve WLAN coverage with no single point of failure. In such networks, 802.11 access points relay information from one to another, hop by hop, in a router-like fashion. As you add users and access points, you add capacity. In addition to 802.11s, other IEEE WGs are currently working to provide mesh networking extensions to their standards, e.g., 802.15.5, 802.16a and 802.20.

This section provides an overview of the mesh networking technology. In particular, starting from commercial case studies, we describe the core building blocks and distinct features the wireless mesh networks should be based on. We provide a survey of the current state of the art for off-the-shelf and proprietary solutions to build wireless mesh networks. Finally, we address the challenges of designing a high-performance, scalable and cost-effective wireless mesh network.

2.2. Popular Commercial Applications for Wireless Mesh Networks

Several emerging and commercially interesting applications for commodity networks have been recently deployed, which are based on the wireless mesh network architecture. To identify all the possible applications exploiting the mesh networking paradigm would be too ambitious for the scope of this survey. Consequently, in this section we focus on providing "case studies" that benefit of wireless mesh networks, i.e., concrete and operating implementations of mesh networking, which exemplify the potentialities behind this radically new framework.

2.2.1. Intelligent Transportation Systems

Several public transportation companies, government agencies, and research organizations are looking for viable solutions to realize *Intelligent Transport Systems*, i.e., integrated public transportation systems that are built to be safe, cost-effective, efficient and secure. The wireless mesh could be the flexible solution to implement the information delivery system required to control the transportation services as depicted in Figure 2.1(a). An example for this application scenario is the Portsmouth Real-Time Travel Information System (PORTAL), a system, part of a citywide public transportation communications network, aimed at providing real-time travel information to passengers[3]. This system is realized by equipping more than 300 buses with mesh technology provided by MeshNetworks Inc. The wireless mesh network allows anybody to display, at more than 40 locations throughout the city, real-time information on the transportation services, like where his or her bus is, which its ultimate destination is, and when it's scheduled to arrive. The same system is also expected to be used to address and alleviate transportation congestion problems, for pollution control and to improve transportation safety and security.

2.2.2. Public Safety

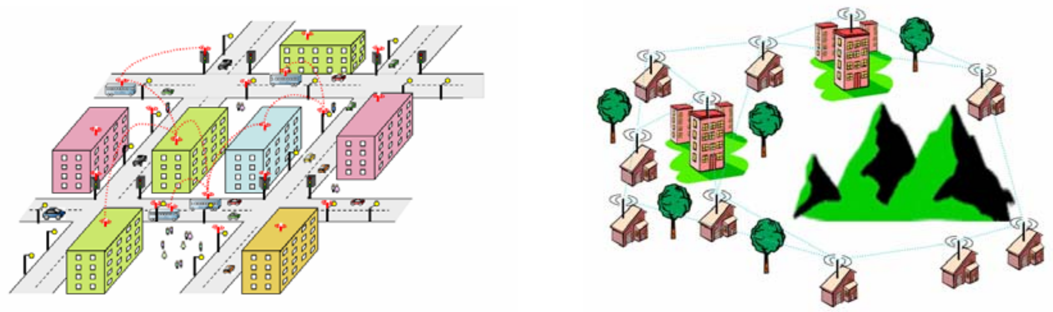
The 9-11 events have dramatically increased the interest in *Public Safety* (police, fire departments, first responders and emergency services), creating additional demand and urgency for wireless network connectivity providing mobility support, reliability, flexibility and high bandwidth. For years, solutions based on cellular technologies have been used, but

[3] http://www.portsmouth.gov.uk/.

they have proved to be unsatisfactory in many aspects. Particularly, cellular data networks promise a near-ubiquitous coverage and allow high-mobility speeds, but data rate is limited, even lower than a typical dial-up connection, and the network infrastructure is extremely costly. Wireless mesh networks appear to be the natural solution to address the needs of law enforcement agencies and city governments. Currently, several mesh networks are operating to provide public safety applications. For instance, the San Matteo Police Department in the San Francisco Bay Area has equipped all its patrol cars with laptops, and motorcycle and bicycle patrols with PDAs, employing standard 802.11b/g wireless card for the communications. The outdoor wireless network is built by using mesh networking technology provided by Tropos Networks. Particularly, more than 30 Tropos Wi-Fi access points were installed throughout downtown to provide ubiquitous coverage to the zone. Tropos proprietary software components are installed over the access points, providing self-discovery and self-configuring functionalities, communications privacy, and centralized network management and control.

2.2.3. Public Internet Access

Internet Service Providers (ISPs) are strongly looking for integrated solutions to implement *Public Internet Access*, which could simultaneously target the market of residential, business and travelers. A growing number of both small and big ISPs are deploying solutions based on Wi-Fi technologies to provide broadband wireless Internet access. The wireless mesh networks are the ideal solution to provide both indoor and outdoor broadband wireless connectivity in urban, sub-urban and rural environments without the need of extremely costly wired network infrastructure. An example of this is the metro-scale broadband city network activated on April 2004 in the City of Cerritos, California, and operated by the Aiirmesh Communications Inc., a Wireless ISP (WISP) company. This network is build up with Tropos-based mesh technology and covers a city area as large as eight square miles using more than 130 outdoor access points, less than 20% of them directly connected to a wired backhaul network. This significant reduction of network installation costs ensures to rapidly deploy a metropolitan broadband network that is cost effective even with a limited potential subscriber base as in rural or scarcely populated urban areas (as depicted in Figure 2.1(b)).



(a) Intelligent Transportation System (ITS)

(b) Residential broadband access too hard to reach and/or scarcely populated areas

Figure 2.1: Emerging commercial applications for wireless mesh networks.

2.3. System and Network Architectures for Wireless Mesh Networks

Wireless meshing has been envisioned as the economically viable networking paradigm to build up broadband and large-scale wireless commodity networks [3]. In this section we will extensively elaborate on this vision to identify the unique and distinct characteristics of this new network architecture.

Several “flavors” of mesh network architectures have been conceived both by industries and academia. However, core building blocks and distinct features may be easily identified in the mesh architecture. A wireless mesh networks is a *fully-wireless* network that employs *multi-hop* communications to forward traffic in route to and from wired Internet entry points. Differently from *flat* ad hoc networks, a mesh network introduces a *hierarchy* in the network architecture, with the implementation of dedicated nodes (called *wireless routers*) communicating among each other and providing wireless transport services to data traveling from users to either other users or access points (access points are special wireless routers with a high-bandwidth wired connection to the Internet backbone). The network of wireless routers forms a *wireless backbone* (tightly integrated into the mesh network), which provides multi-hop connectivity between nomadic users and wired gateways. The meshing among wireless routers and access points creates a wireless *backhaul* communication system, which provides each mobile user with a low-cost, high-bandwidth and seamless multi-hop interconnection service with a limited number of Internet entry points, and with other wireless mobile users. Roughly and generally speaking, backhaul is used to indicate the service of forwarding traffic from the originator node to an access point from which it can be distributed over an external network. Specifically, in the mesh case, the traffic is originated in the users' devices, traverses the wireless backbone and it is distributed over the Internet network. To summarize, Figure 2.2 illustrates the mesh network architecture, highlighting the different components and the system layers.

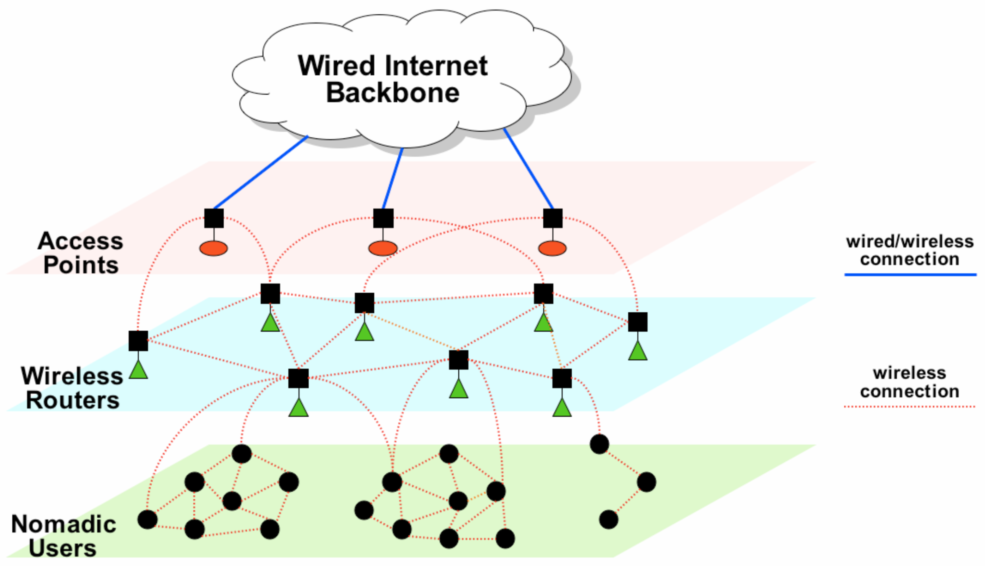


Figure 2.2: A three-tier architecture for wireless mesh networks.

The mesh network architecture addresses the emerging market requirements for building wireless networks that are highly scalable and cost-effective, offering a solution for the easy deployment of high-speed ubiquitous wireless Internet. In the remaining of this section we will further elaborate on the major noticeable benefits of wireless mesh networks that provide substantial arguments in favor of the above claim. The following is not necessarily an exhaustive list of all the possible benefits, but represents an extensive discussion on the motivations behind the mesh networking vision. The interested reader could refer to the *Microsoft Mesh Networking Summit 2004*[4] for a thorough discussion on mesh networking benefits and challenges.

[4] Talks, videos and presentations are online available at http://research.microsoft.com/meshsummit.

Reduction of installation costs. Currently, one of the major efforts to provide wireless Internet beyond the boundaries of indoor WLANs is through the deployment of Wi-Fi hot spots. Basically, the hot spot is an area that is served by a single WLAN, or a network of WLANs, where the wireless clients access the Internet through the 802.11-based access point. To ensure an almost ubiquitous coverage in a metro-scale area, it is needed to deploy a large number of hotspots, due to the limited distance covered by the 802.11 signal. The downside of this solution is an unacceptable increase in the infrastructure costs, because a cabled connection to the wired backbone is needed for every access point. Installing the necessary cabling infrastructure not only slows down the hot spot implementation, but it also significantly increases the installation costs. As a consequence, the hot spot architecture is costly, unscalable and slow to deploy. On the other hand, building a mesh wireless backbone enormously reduces the infrastructural costs because the mesh network needs only a few points of connection to the wired backbone.

Large-scale deployment. In recently standardized WLAN technologies (i.e., 802.11*a*, and 802.11*g*), the increase in data rates has been achieved by using more spectrally efficient modulation schemes. However, for a specific transmit power, shifting towards more efficient modulation techniques reduces the coverage, i.e., the further from the access point the lower the data rate available. Moreover, for a fixed total coverage area, more access points should be installed to cover small-size (e.g., pico) cells [3]. Obviously, this pico-cellularization of WLANs further hinders the scalability of this technology, especially in outdoor environments. On the other hand multi-hop communications offers long distance communications via hopping through intermediate nodes. Since intermediate links are short, these transmissions could be at high data rate, resulting in an increased throughput compared to direct communications. Moreover, the wireless backbone can take advantage of non-mobile, powered wireless routers to implement more sophisticated and resource-demanding transmission techniques than the ones implemented in the users' devices. Consequently, the wireless backbone can realize a high degree of spatial reuse, and wireless links covering longer distance at higher speed than conventional WLAN technologies.

Reliability. The wireless backbone provides redundant paths between each pair of end-points, significantly increasing the communications reliability, eliminating single points of failures and potential bottleneck links within the mesh. The network resilience and robustness against potential problems (as node failures, and path failures due to temporary obstacles or external radio interferences) is also ensured by the existence of multiple possible destinations (i.e., any of the egress points towards the wired Internet) and alternative routes to these destinations.

Self-management. The adoption of peer-to-peer networking to build a wireless distribution system provides all the advantages of ad hoc networking as self-configuration and self-healingness. Consequently, network setup is automatic and transparent to the users. For instance, when adding additional nodes in the mesh, these nodes use their meshing functionalities to automatically discover all possible wireless routers and determine the optimal paths to the wired network. In addition, the existing wireless routers reorganize, taking into account the new available routes. Thus, the network can be easily expanded, because the network self-reconfigures to assimilate the new elements.

2.4. Off-the-Shelf Solutions for Building Mesh Networks

Among the commercial application case studies for wireless mesh networks described in Section 2.2, we omitted the independent (i.e., not owned by ISPs) community networks case. Community networks are systems that allow neighbors to connect their "home networks" together. The advantages of building up community networks are several. For instance, community networks could be used to provide a shared, cost-effective broadband Internet access to a neighborhood, to implement neighborhood surveillance, emergency response systems, and to distribute content that are useful for the neighborhood (e.g., a neighborhood "portal" providing a community with an online bulletin board that allows neighbors to post "for sale" items or trade gossip). Again, wireless mesh could be the technological driver to realize this vision. Nevertheless, the commercial deployment of community networks is still in its infancy. Nowadays, the majority of community network implementations are experimental and non-commercial trials funded and operated by government agencies, non-profit organizations, municipalities and research institutions, and are based on *non-proprietary*, *off-the-shelf* technologies. In this section we will briefly sketch the design choices of one of these experimental trials, the Roofnet network[5], because it well exemplifies the typical advantages and limitations of off-the-shelf solutions for building wireless mesh networks.

Roofnet is an experimental and independent multi-hop 802.11*b* mesh network consisting of about 50 houses located in Cambrige, MA, which has been installed and operated by MIT. The network participants are volunteers that accept to host in their apartments the equipment required to implement a mesh node. One of the main objectives pursued during the design of the Roofnet network has been to employ only open source software and to maintain the costs reasonable low. Consequently, IEEE 802.11 is the radio technology used in the Roofnet community, because cheap network cards operating in unlicensed bands are available. Moreover, many commercial mesh networks rely on directional antennas for increased range, but Roofnet nodes use mainly omni-directional antennas to reduce the per-node costs. Only the gateways, i.e., the nodes bridging the mesh network with the wired Internet backbone, are equipped with directional antennas to provide extended coverage. The user Roofnet node is a computer working as a wireless router, equipped with open source software. Both a wireless and wired network card are mounted on the Roofnet node. The wireless network card is used to connect to the other mesh nodes. A multi-hop routing protocol optimized to find out paths with links of good quality, is used to route the traffic within the mesh. Each Roofnet node also runs a Web server, a NAT, and a DHCP server on its wired Ethernet port. The DHCP server and NAT provide a dynamic host configuration for the user's other computers attached to the home wired LAN. Hence, the Roofnet node acts also as a router for the user's home network. Finally, the Web server provides a simple configuration interface (to turn on and off DHCP, and to set the IP address of the wired interface), a status monitor showing what routes are available and their current metrics.

2.5. Proprietary Solutions for Building Mesh Networks

The growing interest in wireless mesh applications boosted the industrial efforts to develop solutions to make the wireless mesh networks a reality. Several companies and manufactures are now selling proprietary solutions both for indoor and outdoor environments. These solutions adopt radically different approaches and protocols, making these systems incompatible. Some vendors like Tropos, BelAir, Firetide, LocustWorld and Strix, have initially focused on products based on standard IEEE 802.11 technologies, but adopting proprietary software solutions. For instance, Tropos' outdoor systems are cellular Wi-Fi network, where each Wi-Fi cell behaves as a wireless routed LAN. The company has developed its own wireless routing protocol, called Predictive Wireless Routing Protocol (PWRP), which doesn't rely only on hop count to detect transmission paths, but compares packet error rates and other network conditions to determine a best path at a given moment. BelAir, Firetide, Tropos and Strix have also 802.11 products for indoor environments, but they adopt radically different solutions. Firetide, Tropos and Strix, for instance sell indoor meshed networks. In the case of Firetide's and Tropos' products, their outdoor and indoor access points provide the same functionalities and differentiate mainly in the hardware capabilities (e.g., antenna technologies, power requirements, etc.). BelAir's solutions provide indoor coverage from outdoors, by deploying outdoor devices within line of sight or near line

[5] http://www.pdos.lcs.mit.edu/roofnet/.

of sight of a building, which generate radio signals that penetrate the building windows to illuminate the interior. A special case is the LocustWorld company that produces mesh routers, called MeshBox, which are based on open-source software components. Specifically, the core LocustWorld MeshAP device, which adopts as its routing algorithm a Linux-based implementation of the AODV protocol (a public domain protocol developed by the IETF MANET Working Group), is available for download from the LocustWorld website as an open product. On the other hand, commercial projects are required to pay for fully assembled MeshAPs, hardware components and customized functionalities.

Several other vendors like Radiant and MeshNetwork are manufacturing solutions based on proprietary radio technologies. The motivation behind this design choice is that the 802.11 technology has been developed to provide very high data rates over short distances to stationary computers using a very low cost, low powered radio. Consequently, the 802.11 radio technology is not optimized to support mobile and wide range applications. For this reason, the MeshNetwork company has developed a proprietary radio platform, called Quadrature Division Multiple Access (QDMA™), which includes capabilities such as multi-tap rake receivers (commonly found in cell phones) and real-time equalization algorithms to compensate for the rapidly varying RF conditions typically encountered in real-world mobile environments. The MeshNetwork company has also developed a proprietary hybrid ad hoc routing protocol that combines both proactive and reactive routing algorithms, called MeshNetworks Scalable Routing (MSR™) protocol. The MeshNetworks' radio technology still operate in the ISM unlicensed band (2.4 GHz). Other vendors like Radiant Networks, have developed proprietary radio technologies working in licensed bands in the range 26/28 GHz.

2.6. Open Standards Implementing Wireless Mesh Networking Techniques

The open standard radio technologies are essential for the industry because they enable economies of scale, which bring down the cost of equipment and ensure interoperability. For this reason several IEEE standard groups are actively working to define specifications for the wireless mesh networking techniques. These standardization activities differ in the network types they are targeting. In particular, special task groups have been established to define the requirements for mesh networking in Wireless Personal (WPAN), Local (WLAN) and Metropolitan (WMAN) Area Networks. Although with different degrees of maturity, the following emerging standards may be identified: IEEE 802.11s, IEEE 802.15.5, IEEE 802.16a and IEEE 802.20. This section is not aimed at providing a detailed description of these proposed specifications, but at shedding light on the different efforts currently ongoing to implement mesh networking features in the future wireless technologies[6].

IEEE 802.15.5. The IEEE 802.15 project is devoted to the definition of PHY and MAC specifications for establishing short-range wireless connectivity for small groups of fixed, portable and moving computing devices, such as PCs, Personal Digital Assistants (PDAs), peripherals, cell phones, pagers, and consumer electronics. On November 2003, the IEEE P802.15.5 Mesh Network Task Group was formed to determine the necessary mechanisms that must be present in the PHY and MAC layers of WPANs to enable mesh networking. The use of mesh networking in the WPAN environment is motivated by considering the power limitations of the mobile devices. Specifically, employing mesh-like multi-hopping communications increases the coverage of WPANs and allows using shorter links, providing both higher throughputs and fewer retransmissions. Indeed, meshing capabilities are particularly important when using UWB communications, because the bandwidth of UWB

[6] Draft standards and public documentation can be downloaded from the IEEE 802 Standards Committee web site (http://www.ieee802.org/).

wireless links decreases very rapidly (the indoor channel rolls off as the third power of distance). In this case, using shorter links significantly increases the throughput. However, the challenge is to integrate the mesh networking paradigm into 802.15-like MAC protocols. In particular, the 802.15.1 MAC adopts a cluster-based network architecture, where devices are grouped in small "piconets", each with a piconet controller. Moreover, considering the limited resources available in these digital devices, a lightweight implementation of mesh networking techniques should be devised.

IEEE 802.11s. The IEEE 802.11 Working Group is an umbrella that contains several standard committees that are developing technologies for the WLAN environment. The efforts of the currently undergoing standardization activities promise to lead in the near future to the availability of highly interoperable 802.11-based standards providing higher speeds (more than 100 Mbps), QoS support, faster handoffs, and several additional capabilities. Relevant to the mesh networking paradigm is the extension under development by the P802.11s *ESS Mesh Networking* Task Group. The scope of this TG is to extend the IEEE 802.11 architecture and protocol for providing the functionality of an *Extended Service Set* (*ESS*) *Mesh*, i.e., access points capable of establishing wireless links among each other that enable automatic topology learning and dynamic path configuration. The idea behind this proposed amendment is to extend the IEEE 802.11 MAC protocol to create an IEEE 802.11 Wireless Distribution System that supports both broadcast/multicast and unicast delivery at the MAC layer using radio-aware metrics over self-configuring multi-hop topologies. The 802.11s TG is expected to start discussing on proposals for the standard specification in the second quarter of 2005; however the release of the completed standard is not expected before the end of 2006.

IEEE 802.16a. In 1999, the 802.16 Working Group has been established to address the "first-mile/last-mile" connection in Wireless Metropolitan Area Networks (WMANs), working towards LMDS-type architectures for broadband wireless access. The WirelessMAN network, as specified in the 802.16 standard [4], employs a point-to-multipoint (PMP) architecture where each base station (BS) serves a number of subscriber stations (SSs) in a particular area. A PMP system is a star-shaped network where each subscriber connects to the same central hub. The BS transmits on a broadcast channel to all the SSs, while the SSs have point-to-point links with the BS. At the high frequencies (>10 GHz) used in 802.16 systems, line-of-sight (LOS) communications are needed because the system can tolerate a limited amount of multi-path interference. The need of reliable non-line-of-sight (NLOS) operations, together with the opportunity of expanding the system scope to license-exempt bands, has led to the development of the IEEE 802.16a standard. The adoption of NLOS operations allowed including in the 802.16a standard mesh extensions. It is useful to consider how the TDMA-based MAC layer of 802.16a systems supports this optional Mesh mode. In Mesh mode all the SSs may have direct links with other SSs, and the data traffic can be routed through other SSs and can occur directly between SSs. Communications in the direct links can be controlled by either a centralized or distributed algorithm. In the centralized scheduling, the BS determines the flow assignment from the resource requests of the SSs. Subsequently, the SSs determine the actual schedule for their neighbors (i.e., the SSs which have direct links with) from these flow assignments by using a common algorithm. In the distributed scheduling, all the nodes including the BS shall coordinate their transmissions in their two hop neighborhood and shall broadcast their schedules (available resources, requests and grants) to all their neighbors. Although the definitive standards have been already released, commercial products compliant to them are just appearing on the market. For this reason, the WiMAX forum has been established, which is working to facilitate the deployment of broadband wireless networks based on the 802.16 suite of standards by promoting and ensuring the interoperability of manufactured equipments (similarly to what the Wi-Fi Alliance did in promoting the IEEE 802.11 standard for wireless LANs).

IEEE 802.20. Recently, several IEEE working groups are turning their attention to mobile broadband. On December 2002, the establishment of IEEE 802.20, the Mobile Broadband Wireless Access (MBWA) Working Group, was approved. 802.20 systems are intended to provide ubiquitous mobile broadband wireless access in a cellular architecture (e.g. macro/micro/pico cells), supporting the mesh networking paradigm (i.e., NLOS communications) both in indoor and outdoor scenarios. Simultaneously, the IEEE 802.16 WG, under the Task Group *e*, is developing an amendment to the 802.16a specification to support subscriber stations moving at vehicular speeds, conceiving a system for combined fixed and mobile broadband wireless access. Despite the fact that 802.16*e* and 802.20 standards will both specify new mobile air interfaces for wireless and mobile broadband services, there are some important differences between them. 802.16*e* will add mobility in the 2 GHz to 6 GHz licensed bands, while 802.20 aims for operation in licensed bands below 3.5 GHz. Moreover, 802.16*e* is looking at the mobile user walking around with a PDA or laptop, while 802.20 addresses high-speed mobility issues (speeds up to 250 kilometers per hour). More importantly, the 802.16*e* specification will be based on an existing standard (802.16*a*), while 802.20 is starting from scratch. Both the working groups are still in a preliminary stage and no public specifications have been released yet. The 802.20 project plans to release a draft standard to submit for approval in the second semester of 2006.

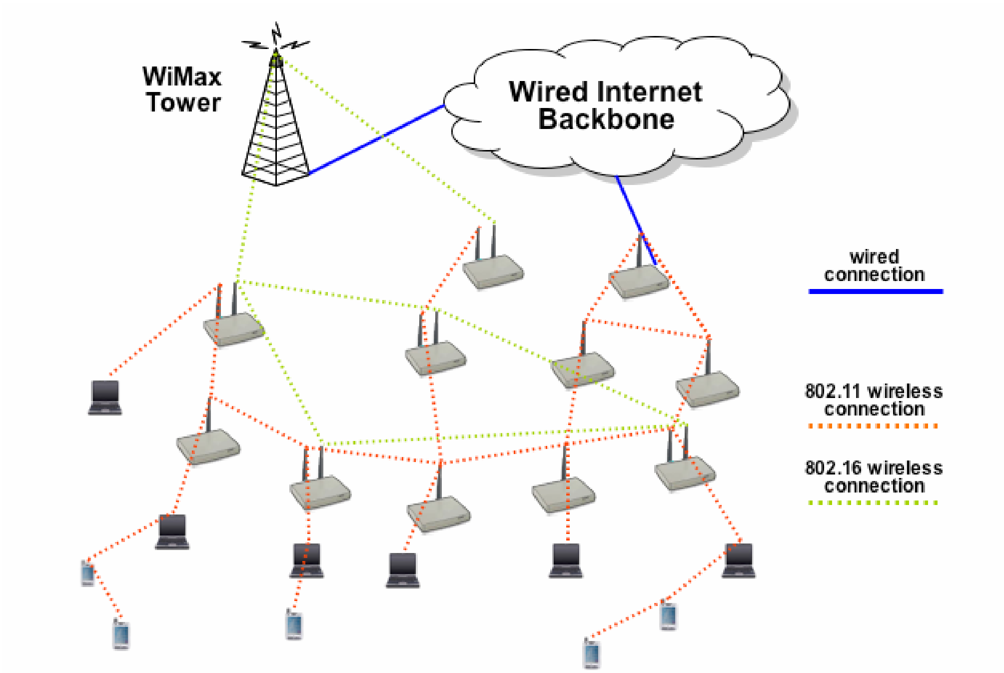


Figure 2.3: Integration of WiMAX and Wi-Fi technologies in large-scale wireless mesh networks.

Although the above standards target different network environments, these technologies are not complementary and overlap as far as many proposed functionalities. Consequently, networks operators that want to deploy solutions for the last-mile broadband wireless Internet access can take advantage, for instance, of both the emerging 802.11s and 802.11a products. Nowadays, Wi-Fi-based solutions appear advantageous because they are already established, and they operate in unlicensed cost-free frequency bands. Nevertheless, it is feasible to envision integration between WiMAX and Wi-Fi, particularly considering that the 802.16a MAC and PHY layers are optimized for long-distance wireless links. In Figure 2.3 we show a wireless mesh that fully exploits the advantages of the WiMAX technology, implementing both wireless Point-to-Multipoint (PMP) communications between the wireless routers and the Internet backbone, and mesh-based communications among the wireless routers. Once the wireless mesh networks based on Wi-Fi products is installed, the integration of WiMAX will be straightforward. Indeed, 802.16 wireless links can be easily added to the existing network to either expand the networks or to introduce additional capacity in the wireless backbone. Consequently, WiMAX products can offer low-cost, flexible alternatives to build the wireless backbone in the outdoor scenarios.

2.7. Key Research Challenges

The mesh network architecture, as conceived in Section 2.3, is an economically viable solution for the wide deployment of high-speed, scalable and ubiquitous wireless Internet services. However, the major technical challenges of building a large-scale and high-performance multi-hop wireless backhaul system are not solved yet. Indeed, the wireless infrastructure meshing formed through multi-hop communications among wireless routers and access points (as depicted in Figure 2.2) cannot be simply treated as a large multi-hop ad hoc network, because the structure and the functionalities of such network are radically different from the ones of a *general* ad hoc network. In practice, this simplification will undoubtedly lead to the well-known scalability limits of ad hoc networks, due to the dramatic degradation of throughput and delay performance as the network diameter increases [5]. Consequently, one of the major problems to address while building a multi-hop wireless backhaul network is the *scalability* of both the network architecture and protocols. Hence, in the following sections we will discuss the most relevant and promising research activities, focusing on the design and development of a scalable and high-performance wireless backbone for mesh networks.

2.7.1. High Capacity and Reliable Radio Interfaces for the Wireless Backbone

Currently, there are several research efforts to improve the *capacity* of wireless mesh networks by exploiting alternative approaches as multiple radio interfaces, MIMO techniques, beam-forming antennas, opportunistic channel selection[7]. Multiple channels and/or radio interfaces could increase network capacity by exploiting the independent fading across different frequencies or the orthogonality of frequency bands. Similarly, systems employing multiple antennas for both transmitting and receiving (generally called Multiple-Input-Multiple-Output systems) improve the capacity and reliability of wireless backbones by exploiting antenna diversity and spatial multiplexing. Diversity provides the receiver with several (ideally independent) replicas of the transmitted signal and is therefore a powerful technique to combat fading and interference. On the other side, spatial multiplexing divides the channel into multiple "spatial channels" through which independent data streams or signals can be coded and transmitted simultaneously. As a consequence, diversity techniques make the channel a less fading one, which is of fundamental importance for wireless backbones, where deep fades can occur and the channel is changing slowly, causing the fades to persist over a long period of time. Nevertheless, when strong interference is also present, diversity processing alone cannot improve the signal. To cope with interference, smart antennas or adaptive array processing can be utilized for enhancing both the energy efficiency and the multiple-access interference rejection capability of the high-throughput wireless backbone. The key idea is to exploit the beamforming capability of the transmit/receive antenna arrays. Roughly speaking, beamforming creates an effective antenna pattern at the receiver with high gain in the direction of the desired signal and low gain in all other directions. Hence, the exploitation of directional transmissions could suffice to ensure a wireless backbone with high speed and high degree of spatial reuse [6].

[7] See, for example, the presentations held at the Microsoft Mesh Networking Summit 2004 (available online at http://research.microsoft.com/meshsummit).

2.7.2. Designing Scalable and Opportunistic Networking Functions

Although the use of multiple antennas at the wireless routers in combination with signal processing and coding is a promising mean to provide a high capacity wireless backhaul system, it is not enough alone to achieve a scalable wireless backbone. For instance, it is well-known that as the number of users increases random medium access control (MAC) protocols suffer for increased contention in the network. Moreover, the users' traffic traversing the wireless backbone does not have a unique, fixed destination, but rather can be delivered to any wired-access point. In addition, several paths may contemporarily exist to reach a given access point; paths capacity and channel bandwidth could be highly variable. Consequently, new scalable and distributed scheduling, MAC and routing protocols have to be designed to efficiently manage data traffic. These algorithms must be aware of the characteristics of the physical channel and this leads to the need for a *cross-layer* design among physical and networking functions. Nearly all the literature focusing on cross-layering to optimize networking functions exploits the *multi-user* diversity, i.e., the condition when in a system with many users, different users experience peaks in their channel quality at different time instants [7]. For instance, in this case it is proved that the scheduler should allocate transmission opportunities to users with the most favourable channel conditions [7]. The mesh network environment adds further degrees of freedom in the scheduling process, because the scheduling policies could exploit additional types of diversity such as spatial diversity (spatial channels opened by multi-antenna wireless backbone implemented at physical layer) and frequency diversity (radio technologies using multiple frequency channels) to enhance throughput. Moreover, the design of scheduling policies for a multi-channel, multi-hop and multi-destination system is extremely challenging because the opportunistic selection of the high-quality channel cannot be performed locally in the single wireless router, but should be coordinated among all the wireless routers forming the backbone network. Consequently, the scheduling process in a wireless mesh network is intrinsically distributed, where the coordination among wireless routers is achieved via the exchange of messages containing information on channel conditions and traffic demands.

The MAC and PHY layers play a crucial role in providing the scalability and performance optimization required by wireless mesh networking. Furthermore, to fully exploit the potential capacity improvement ensured by the adoption of optimized transmission and antenna technologies, it is fundamental that the routing protocol discovers high-quality routes by explicitly considering the current network conditions. Most of the current routing protocols for multi-hop communications typically choose optimized (in the sense of minimum hop-count, maximum lifetime of the route, or maximal residual power in the nodes along the route) paths without taking into account the link quality. Therefore, several research efforts are devoted to the definition of novel routing metrics that correctly account for the loss rate and channel bandwidth of each link forming the path [8]. Moreover, the routing protocol for a wireless backbone needs to be redesigned not only to deal with the path diversity, but also to address the distinct nature of the wireless backbone network with respect to a general ad hoc network. In particular, the user traffic to the Internet does not need to follow the same path, but could be forwarded to any of the Internet egress points in the multi-hop wireless backhaul network. Consequently, the routing protocol should opportunistically select the "best wire", i.e., the optimized path, subject to performance constraints, towards any of the wired access points. Finally, the routing protocol could effectively benefit from the existence of non-mobile, powered wireless infrastructure to exploit hybrid ad hoc routing that combines both proactive and reactive techniques. In the wireless backbone, formed by stationary wireless routers, it is reasonable to envision that link-state routing protocol analogous to traditional wired routing protocol as OSPF could be used. Recently, a scalable link state routing protocol has been designed that minimizes the cost of maintaining a consistent view of the network, called Hazy Sighted Link State (HSLS) routing [9]. An open source implementation of the HSLS protocol is under development by the CUWiN project.

Finally, it is worth pointing out that mesh networking, as a special case of ad hoc networking, should fully implement self-management, self-configuration and self-healing features in all the layers of the network architecture. Consequently, a key research challenge is also to ensure that the scalable and opportunistic networking functions designed specifically for the mesh networks effectively fulfill the requirements of the peer-to-peer networking paradigm adopted in the wireless backbone.

2.7.3. System-Wide Resource Management

The wireless backbone forming the core of a mesh network provides a backhaul communication service. The end users' traffic is transparently routed to and from the wired Internet employing multi-hop wireless path traversing the wireless backbone. It is an essential requirement for the backhaul network, to ensure that all the users in the network achieve a *fair* share of the system resources. Unfortunately, current networking protocols are not appropriate for multi-hop wireless backbone networks, usually inducing severe unfairness and scarce performance to user located far from the available Internet egress points. Hence, it is needed to develop a coordinated multi-hop resource management algorithm to achieve high performance while preserving a system-wide notion of fairness [2]. Fairness in ad hoc networks has been extensively studied in the last years. However, the distinct structure of the wireless backbone requires defining a new fairness model to address the distinct objectives and characteristics of such network. In particular, both max-min and proportional per-flow fairness are inadequate fairness objectives in the case of multi-hop wireless backhaul networks, because wireless routers must manage the aggregated traffic flows traversing the network. In [10] a novel fairness model has been proposed that addresses the requirements of multi-hop aggregated flows, aiming at eliminating the spatial bias, by ensuring that each user will receive the same fair share of resources independently of how far it is from the Internet entry point, i.e., independently of its spatial location.

A coordinated resource management is required not only to tackle the issue of providing system-wide fairness and to exploit the spatial reuse in the wireless backbone, but also to provide a prompt reaction to the variations in system capacity due to changes in traffic patterns, channel conditions and contention. Considering the intrinsic large-scale nature of the wireless backbone to achieve system-wide performance objectives the resource management algorithm must be distributed. However, a careful design approach of the network control has to be employed to trade-off the additional overhead of increased protocol information required to perform a more precise control and the benefits deriving by the opportunistic exploitation of this information. Consequently, the analysis of system capacity and scalability should incorporate the impact of protocol overheads and operations.

2.8. References

[1] I. Chlamtac, M. Conti, J.J.-N. Liu, "Mobile ad hoc networking: imperatives and challenges", *Ad Hoc Networks*, vol. 1, no. 1, Jul. 2003, pages: 13-64.

[2] R. Karrer, A. Sabharwal, E. Knightly, "Enabling Large-Scale Wireless Broadband: The Case for TAPs", *ACM SIGCOMM Computer Communication Review,* vol. 34, no. 1, Jan. 2004, pages: 27-34.

[3] E. Baccarelli, M. Biagi, R. Bruno, M. Conti, E. Gregori, "Broadband Wireless Access Networks: a Roadmap on Emerging Trends and Standards" in *Broadband Services: Business Models and Technologies for Community Networks*, I. Chlamtac, A. Gumaste, C. Szabo (Editors), *John Wiley & Sons* (Publisher), Mar. 2005.

[4] C. Eklund, R.B. Marks, K.L. Stanwood, S. Wand, "IEEE Standard 802.16: A Technical Overview of the WirelessMAN™ Air Interface for Broadband Wireless Access", *IEEE Communications Magazine*, vol. 40, no. 6, Jun. 2002, pages: 98-107.

[5] P. Gupta, P.R. Kumar, "The Capacity of Wireless Networks", *IEEE Transactions on Information Theory*, vol. 46, no. 2, March 2000, pages: 388-404.

[6] H Bölcskei, A.J. Paulraj, K.V.S. Hari, R.U. Nabar, W.W. Lu, “Fixed Broadband Wireless Access: State of the Art, Challenges, and Future Directions”, *IEEE Communications Magazine*, vol. 39, no. 1, Jan. 2001, pages: 100-108.

[7] X. Quin, R. Berry, “Exploiting multiuser diversity for medium access control in wireless networks”, in *Proceedings* of *IEEE Infocom 2003*, San Francisco, CA, 30 March-3 April 2003, vol. 2, pages: 1084-1094.

[8] R. Draves, J. Padhye, B. Zill, “Routing in Multi-Radio, Multi-Hop Wireless Mesh Networks”, in *Proceedings* of *ACM MobiCom’04*, Philadelphia, PE, Sept. 26-Oct. 1, 2004, pages: 114-128.

[9] C.A. Santivanez, B. McDonald, I. Stavrakakis, R. Ramanathan, “On the scalability of ad hoc routing protocols”, in *Proceedings* of *IEEE Infocom 2002*, New York, NY, June 23-27 2002, vol. 3, pages: 1688-1697.

[10] V. Gambiroza, B. Sadeghi, E. Knightly, “End-to-End Performance and Fairness in Multihop Wireless Backhaul Networks”, in *Proceedings* of *MobiCom’04*, Philadelphia, PA, Sept. 26-Oct. 1, 2004, pages: 287-301.

3. LONG-TERM FUTURE DIRECTIONS

3.1. Introduction

The previous section envisioned the Mesh Networking paradigm as the most interesting evolution of legacy MANETs in the short to medium term. In this section, we take a broader view, and present the idea of opportunistic networking. We believe that this paradigm has the potentialities to become a key building block of future integrated networks.

Traditionally, node mobility in MANETs has been seen more as a problem to cope with (e.g., recover routes broken because of nodes movement) than as a way of creating connectivity opportunities. In the last part of the MobileMAN project we have started to explore the latter aspect, by characterising the *inter-contact times* and the *contact durations* between users carrying mobile devices. We envision an *opportunistic networking* scenario where mobile devices are not well connected like in traditional MANETs, but they are frequently disconnected, and exploit any contact opportunity with other mobile devices to send/forward information. This is a challenging scenario from a technical standpoint. Moreover, it is very interesting also from an economic standpoint. Applications could be designed that exploit the widespread use of portable mobile devices (e.g., cell phones, PDAs, etc.), without any additional requirements in terms of network planning. The results presented in Section 3 represent a first step in the analysis of this scenario.

In a wider perspective, the opportunistic networking concept calls for heterogeneous integrated networks. The user may get or be in touch with different networks (e.g., cellular, Wi-Fi, legacy MANET, etc.), and would opportunistically choose the most appropriate technology that suits her needs. Therefore, at the end of this section we elaborate on issues related to heterogeneous routing. We claim that allowing for different, co-existing, routing policies in the same network can be of great value to users. Different routing protocols will be tied to different cost/QoS tradeoffs, as well as to different devices' capabilities. Users will specify the QoS they want to receive, and the cost they are willing to pay for (e.g., a broker would be quite willing to pay for real-time dispatch of quotes she is trading, while would prefer a cheap, best effort, delivery of a greeting message for her friend's birthday). To complement the MobileMAN routing protocol suite, with respect to the protocols already described in previous deliverables, we present Landmark Guided Forwarding, which exploits availability of even imprecise localisation information to route information within a MANET. LGF can be a valid option to route information through those nodes that have some localisation capability, such as GPS.

3.2. Frequent Disconnection

We use the term opportunistic networking to refer to data exchanges based on the connection "opportunities" that arise whenever mobile devices happen to come into wireless range due to the mobility of their users. Such situations are prevalent in many regions of the world where broadband access infrastructure has limited coverage (as well as cost and application constraints). Thus, users have "islands of connectivity" (e.g. home or work), but are also likely to sporadically be in range of other users while in between. It is also worth noting that access infrastructure is vulnerable to natural disasters or other failures, and that in such exceptional situations, opportunistic networking may be the only feasible way to carry important data.

Opportunistic networking is a sub-class of both Delay-Tolerant Networking (DTN) [1] and Mobile Ad-Hoc Networking (MANET) [2]. In this section we are concerned specifically with the part of the design space where nodes are often out of contact with one another. This work makes three key contributions.

1. We present the design and results of an experiment which provides a picture of some typical human mobility patterns in the context of opportunistic connections.

2. We analyse both the data gathered by ourselves as well as two other large data sets made available from other mobility experiments and show that an approximate power law holds over an extended range of values for the inter-contact times of nodes in all four experiments.

3. We provide a proof that one major class of previously proposed opportunistic forwarding algorithms will not perform well in these conditions and propose measures which may increase the performance of such algorithms.

3.2.1. Related work

We have investigated three areas for related work:

Measurement

A number of research groups have conducted studies into mobility in the context of networking. Many of these are aimed at analysing and informing the design of infrastructure-based networks, but their data and results are also relevant for opportunistic networks. This category includes Balazinska and Castro’s study [3] as well as the data gathered at UCSD [4] and Dartmouth [5] which we analyse in this section.

Modelling

Much of the work in DTN and MANETs concerns the modeling of mobility or location [6, 7]. The goal of the models has typically been to drive the evaluation of routing schemes which assume that the majority of nodes are connected most of the time. This is not likely to be relevant since we are in the part of the design space where nodes are often disconnected. Indeed, the purpose of this work is to model the distribution of these disconnection times and its impact on forwarding decisions, as noted next.

A common property of many mobility models found in the literature is that the inter-contact distribution decays exponentially over time. One of the simplest examples, introduced in [8], in the case where nodes locations are i.i.d. with a uniform distribution in a bounded region, the success of communication between each sender-receiver pair has a fixed probability $p > 0$ at each time slot, t.

In this situation, if we consider a particular sender-receiver pair, the remaining time to the next contact is distributed exponentially following the distribution

$X:P\ [X > t] = pt$ (1)

This property is also true for the popular random waypoint model, see [9]. In this article, a Brownian motion model is analysed as well. The authors claim that the inter-contact time is in this case stochastically bounded by an exponential random variable.

Forwarding

Su et al.[10] used traces of human mobility patterns gathered with PDAs toevaluate the feasibility of opportunistic networking. Some sensor networks act in an opportunistic fashion, such as Zebranet [11], which uses opportunistic connections between zebra-mounted nodes to transfer sensor data and thus collect statistics about zebra populations. Similar research exists with whales [12]. However, these experiments are largely pragmatic proofs-of-concept.

The most relevant work when trying to find forwarding algorithms for networks that are frequently disconnected, is the algorithm proposed by Grossglauser and Tse in [8], further analysed in [9]. The principal motivation of that work was somewhat different from ours: it was to find the available increase in capacity of the multi-hop radio network as a function of the mobility and the number of nodes. In the process of exploiting mobility and trading it off against transmission, the authors created an opportunistic forwarding algorithm.

3.2.2. Contact Opportunities

Given the nature of intermittent connectivity, it is likely that successful forwarding algorithms are based purely on locally learned information. The regime in some senses is even more resource starved than a MANET where one eschews proactive routing. Thus we need to measure what we can statistically learn locally in a variety of likely scenarios, and then use those measurements to drive the evaluation of appropriate forwarding algorithms.

As envisaged in the introduction, one can imagine that such a regime might operate only in parts of a network, where in other parts, connectivity is maintained. We will discuss this in the conclusions under further work. Gathering data on transfer opportunities In order to conduct informed design of forwarding policies and algorithms for opportunistic networks, it is important to gather data on the frequency and duration of contact between humans (and the devices they carry). However, this is not easy to gather - ideally, a data set would cover a large user base over a large time period, as well as include data on connection opportunities encountered twenty-four hours a day. This presents many practicality issues ; dealing with deployment of mobile devices to a large user population, the battery life of the devices, and minimising the inconvenience to users of carrying the devices so that they are willing to doso at all times.

We first examined the data made available to the community by people who have performed previous measurement exercises. Two data sets emerged, namely from UCSD [4] and Dartmouth [5]. Both make use of WiFi networking, with the former including client-based logs of the visibility of access points (APs), while the latter includes SNMP logs from the access points.

The durations of the logs are three and four months respectively. Since we required data about device-to-device transmission opportunities, the raw data sets were unsuitable for our experiment and required pre-processing. For both data sets, we made the assumption that mobile devices in sight of the same AP would also be able to communicate directly (in ad-hoc mode),and created a list of transmission opportunities by determining, for each pair of nodes, the set of time regions for which they shared at least one AP.

Unfortunately, this assumption introduces inaccuracies. On one hand, it is overly optimistic, since two devices attached to the same access point may still be out of range of each other. On the other hand, the data might omit connection opportunities, since two devices may pass each other at a place where there is no instrumented access point, and this contact would not be logged. Furthermore, it is hard to ensure that the devices are in fact co-located with their owner at all times. Despite these inaccuracies, the WiFi traces are a valuable source of data, since they span many months and include thousands of nodes. In addition, considering two devices connected to the same

AP are potentially in contact is not altogether unreasonable, as these devices could indeed communicate through the AP, without using end-to-end connectivity.

In response to the limitations of the previous traces for our work, we set up our own experiments making use of the iMote platform made by Intel Research. iMotes are derived from the Berkeley Mote3, with the current version based around the Zeevo TC2001P system-on-a-chip providing an ARM7 processor and Bluetooth support. Along with a 950mAh CR2 battery, each iMote was enclosed in packaging designed to be convenient for test subjects to continually carry. Two types of packaging were made available : some iMotes were made into keyfobs while others were enclosed in small boxes. Subjects were asked to pick the form factor which allowed them to conveniently keep the iMote with them at all times, with most simply attaching the iMote to their keys.

The Bluetooth specification indicates ten seconds should be used, our experience with the iMote hardware is that the vast majority of nearby nodes are seen in the first five seconds or not at all. Performing an inquiry is power-intensive and it is important to minimise time spent in this mode. The iMotes spend 120 seconds (plus or minus twelve seconds with a uniform random distribution) in a sleep mode where they are able to respond to inquiries. The reason for introducing randomness is that iMotes cannot respond to inquiry while themselves performing inquiry. Without randomness, iMotes might synchronise in such a way that they would never see each other. The two minute inquiry period was chosen to provide an estimated lifetime of one week for the iMotes with CR2 batteries. During the experiment, each iMote used flash memory to log “contact” data for all visible Bluetooth nodes (including iMotes as well as other Bluetooth devices), with each contact being represented by a tuple (MAC address, start time, end time). A twenty-four hour pilot deployment was performed in order to iron out software bugs and refine the deployment methodology and packaging mechanisms.

Two iMote experiments (“iMote A” and “iMote B”) were conducted. Experiment iMote A included seventeen researchers and interns working at Intel Research Cambridge, while iMote B involved eighteen doctoral students and faculty comprising a research group at the University of Cambridge Computer Lab. Unfortunately, real world factors contributed to the malfunction of some of the iMotes. As a result, the two experiments resulted in data from nine and twelve iMotes.

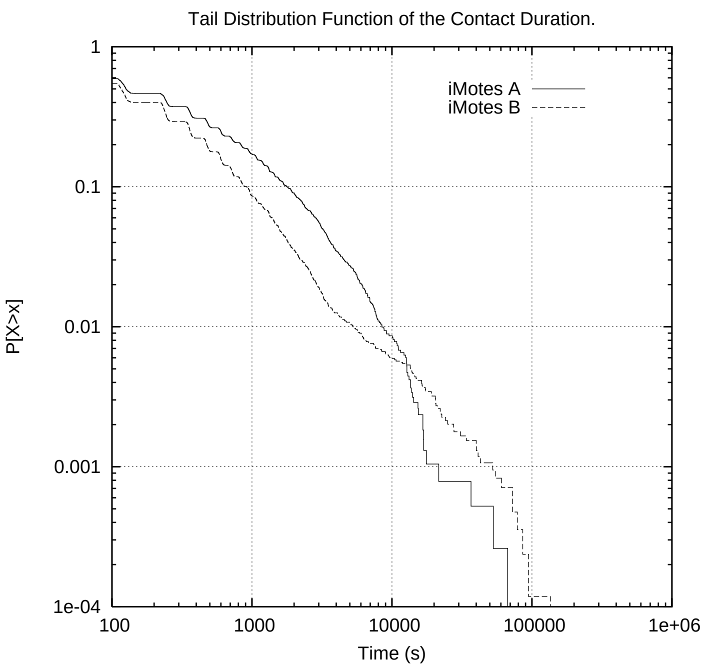


Figure 3.1: Distribution of contact duration for Bluetooth

The iMotes were collected after their battery had expired, and the flash memory was read. A number of post-processing steps were undertaken on the data. A time basis consistent for all iMotes was reconstructed by (1) using the known start time of the experiment, and by using a program to find “mutual” sightings (i.e. where two iMotes see each other) of long duration, and (2) determining the clock offset between the two iMotes. The synchronisation was checked manually. iMotes contacts were classified as “internal” with other iMotes and “external” with other types of device. External contacts are a valuable source of data. iMotes are deployed to a small set of advanced users. The external contacts are numerous and include anyone who has an active Bluetooth device in the vicinity of the iMote users, thereby measuring the actual Bluetooth deployment. The Internal contacts, on the other hand, represent the contacts that each of our participants would have, if they were equipped with devices (such as smart phones) which are always-on and always-carried, which supported opportunistic networking.

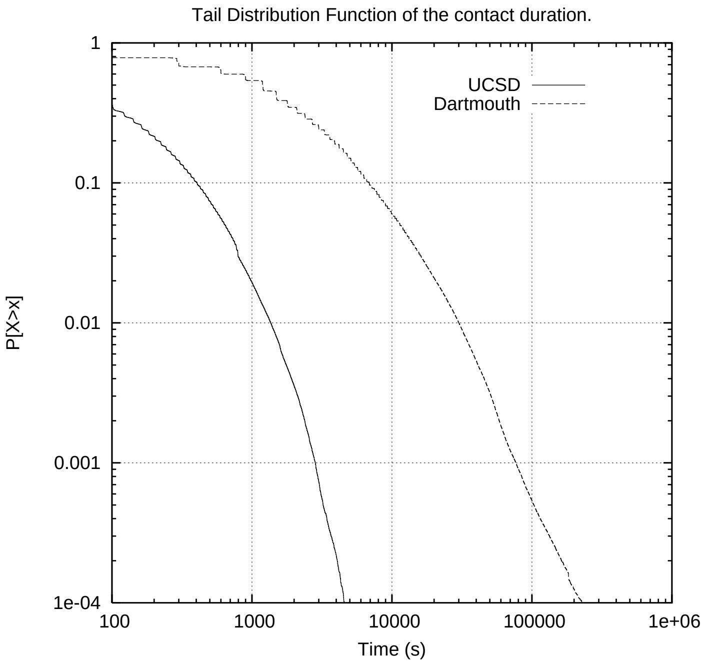


Figure 3.2: Distribution of contact duration for WLAN

The iMote experiments have the advantages that the users are more likely to carry the small-form-factor iMote at all times, and that logging takes place wherever the user is and not just when the users are near APs. The WiFi-based experiments have larger user populations and durations, and include all contacts occurring at the instrumented locations. Thus, the data sets are complementary in many ways:

Characterization of contact opportunities

This section reports our observations on the four mobility data sets described above. The consequences of these observations on opportunistic networking forwarding algorithms will be described in the following section. We are interested in the characteristics of connection opportunities, i.e. how many and when do they occur, how often and how long. We choose to characterize these opportunities in term of contacts.

The contact duration is the time interval for which two network devices can communicate when they come into range. The number of such contacts and the distribution of contact durations is an important factor in determining the capacity of opportunistic networks. It gives insight on how much data can be transferred at each opportunity.

The inter-contact time is the time interval between two contacts. This parameter strongly affects the feasibility of opportunistic networking, and has rarely been studied in the literature. The nature of the distribution will affect the choice of suitable forwarding algorithms to be used to maximize the successful transmission of messages in a bounded delay.

In this work, we focus on the analysis of the distribution of inter-contact time. For completeness, we also briefly discuss the contact time distribution. Two remarks must be made at this point. First, we computed the distributions of these two variables per event (i.e. considering each one of the value taken by these periods of time). We did not compute the distribution of the variable as seen at a random instant in time by the pair. These two distributions are in correspondence by a classical result from renewal theory (see [13]).

In addition to that, our results are influenced by the duration and granularity of the experiments. For event lengths approaching the duration of the experiment there is an artificially

lower likelihood of observation, and events lasting longer than the experiment cannot be observed. Similarly, for short event lengths, the data is affected by the granularity of measurement (which ranges from twenty seconds to five minutes in our data sets).

Inter-contact time

We study inter-contact time first as it is the parameter that has the most significant impact on the feasibility of opportunistic networking. Inter-contact time affects the frequency with which packets can be transferred between networked devices. We plot inter-contact time distributions for all four data sets (i.e. iMote A, iMote B, UCSD, and Dartmouth). The significant region is the middle of the graph, with the leftmost and rightmost parts showing artifacts due to the granularity and duration of the experiments as described above. In this region, all four experiments show the same behavior, an approximate power law, as evidenced by the straightness of the curve. A power law is characterised by it’s coefficient reflecting the slope of the line on log-log graphs. For the iMote experiments, the coefficient is 0.5 for the range [100s; one day], with a slightly convex distribution (flatter than that of a power law). The Dartmouth data exhibits an approximate power law with a coefficient of 0.33 on the range [100s; 1 week]. The UCSD distribution coefficient is also 0.33, but over a more limited range [100s; two days].

The approximate power law shape means that the inter-contact distribution is heavy-tailed over this range - i.e. the tail distribution function decreases slowly. This is contrary to the exponential decay of many mobility models put forward in the literature. As a result, opportunistic networking algorithms which have been designed around exponential models must be re-evaluated in the light of our observations (see next section). In the Bluetooth traces, 15% of inter-contact durations are greater than one hour, and 5% are greater than one day. In the WiFi experiments, the large experiment duration allows us to analyse large inter-contact durations. In the Dartmouth trace, we find that they are far from negligible: 20% are more than a day, 10% are more than a week. In the UCSD trace, 15% are more than a day, and 4% are more than one week.

While the WiFi experiments have longer durations, longer inter-contact times may be affected by the more limited mobility of laptops or PDAs as their users may not carry them all the time. As one might expect, the iMote experiments show lower inter-contact times (as illustrated by the lower coefficient of the approximate power law). This is an encouraging sign for the field of opportunistic networking - with the always-carried, always-on nature of devices such as smart phones, more connection opportunities will be found. What is striking is that the same overall pattern, the approximate power law, seems to apply to both Bluetooth and WiFi data despite the differences in experimental methodology.

Internal vs. external iMote contacts

While the above iMote plots used both internal and external contacts, it is also instructive to look at each of these classes in isolation. We observe that the internal contact plot is almost indistinguishable from the plot containing all contacts while the external contact plot has a very similar distribution, again showing an approximate power law. This indicates that our experimental methodology using iMotes is suitable for studying the behavior of the deployed Bluetooth user base.

Impact of time of day on opportunities

Since human movement patterns have daily cycles, it is possible that the results above arise as an artefact of averaging over entire days, and that during given parts of the day different behaviour is seen. In order to address this possibility, we split the day into 3 hour bins. For each inter-contact gap in the experiments, we add the duration to the bin in which the gap starts.

Contact duration

The contact duration is one of the factors affecting the amount of data that can be transferred between nodes when they come into range. Other relevant factors include relative distance pat-terns, the speed of movement, the discovery latency, and wireless network congestion. While the capacity of an opportunistic contact is obviously an important topic for opportunistic net-working, it is not the focus of this work.

3.2.3. Networking with power law-based opportunities

In this section, we study the impact of heavy tailed inter-contact times on the actual performance and theoretical limits of a general class of opportunistic forwarding algorithms. These algorithms can be characterized as “stateless” in that they do not maintain history data or at-tempt to predict performance in the future. Instead, each node takes advantage of contacts opportunities to forward as many packets as possible while in contact with another node. We start by defining formally the class of algorithms we study.

We are interested in a general class of opportunistic forwarding algorithms that relies on inter-mediate nodes to carry data between a source and a destination that might not be contemporaneously connected. Intermediate nodes are chosen purely based on contact opportunism and not using any stored routing information.

The following two algorithms provide bounds for the class of algorithm described above :

wait-and-forward : The source simply waits until its next contact with the destination to communicate.

flooding : a node forwards all its packets to any node which it encounters, keeping copies for itself

The first algorithm uses minimal resources but can incur very long delays and does not take full advantage of the ad-hoc network capacity. The second algorithm delivers packets with the minimum possible latency, but does not scale well in terms of bandwidth, storage, and battery usage.

In between these two extreme cases, there is a whole set of middle-ground algorithms that use various numbers of intermediate nodes, contacts, and packet duplicates. Starting from the most conservative algorithms (wait-and-forward), we analyse incrementally the capacity of our class of algorithms, in the face of power law behaviour of the inter-contact delay distribution.

We first study the two-hop relaying algorithm introduced by Grossglauser and Tse in [8].The two-hop relaying scheme operates as follows. Time is divided in a sequence of even and odd slots. During every even time slot, packets are sent from sources to relays, with the first available contact chosen as a relay The only exception is if the destination is the first node to come into contact with the source, as the packet is transmitted directly. Otherwise, the relay will keep the packet in memory and the source does not send this packet again. Relays deliver to the destination only; going through a second relay is not permitted.

Each relays maintains one queue per destination. We assume that buffers are unlimited, and that packets in each queue are transmitted to the destination in a first-come first-served fashion, at the next time that the relay encounters the destination, in an odd time slot. As queuing is used in the intermediate nodes, the forwarding process of packets sent by the source to a relay needs to be of lower intensity than the packets sent by this relay to the destination. This is the case in the implementation proposed in [8] and we make the same assumption below.

This algorithm is a good candidate to start our study of the impact of power law inter-contact times on opportunistic forwarding for the following three reasons :

The algorithm was shown to maximize the capacity of dense ad-hoc networks, under the condition that nodes are i.i.d. uniformly in a bounded region.

This result depends strongly on the mobility process of nodes. Authors of [8] assumed an exponential decay of the inter-contact time. The same result has been proven for packets following Brownian motion or random waypoint mobility model [9].

[8, 9] have shown that the packets experience a finite expected delay under these conditions.

3.2.4. Summary, conclusion and future work on Opportunistic Networks

We study an ad-hoc network scenario, called opportunistic networking, where inherent mobility and the occasional connection with other devices are used to transfer data.

We establish a first major result, which is that in four different and independent data sets, the distribution of the “inter-contact time” between nodes in an opportunistic networking environment follows an approximate power law over a large range, with power law coefficient less than one. This result is not consistent with the exponential decay predicted by all existing node mobility models used to date in ad-hoc networking.

We show that a class of stateless forwarding algorithms, that have been proved to deliver packets with a bounded delay in the case of exponential decay inter-contact times, have indeed an infinite expected delay when mobility follows approximate power law inter-contact with coefficient under 1. We prove that using multiple intermediate relays is sufficient for these algorithms to converge when the power law coefficient is located between 1 and 2. Above two, these algorithms converge naturally.

The implications of our work for the research community are as follows: 1. Current mobility models (e.g. random waypoint, uniformly distributed locations) do not have the characteristics observed in our human mobility experiments. New mobility models are therefore required in order to facilitate evaluation of potential opportunistic data transmission schemes. 2. Little work has been done in the area of informed design of opportunistic forwarding algorithms—this remains an area ripe for study. Suitable directions for work might involve the sharing of recent contact information between nodes, leading to a more careful selection of potential relay nodes which are likely to have a short path to the destination, while also being independently moving as compared to other chosen relay nodes. We plan to continue our research in a number of directions. Firstly, we wish to characterise the contacts patterns between nodes as well as the inter-contact patterns, looking at how con-tact duration, node speed, relative distance, discovery latency, and network congestion affect the capacity of an opportunistic contact for Bluetooth and WiFi networks. Secondly, we wish to explore tractable forwarding algorithms for opportunistic networks, taking into account the lessons above of avoiding long-delay paths and of sending redundant copies over independently-moving relays. Finally, we wish to examine a hybrid of opportunistic and infrastructure based networking, and the delay and bandwidth characteristics present in such conditions. This has led to a new EU IST proposal.

3.3. Hybrid Geo/Topo routing and location inaccuracy

Today, MANET schemes are largely purely ad hoc. In the future when the network is part of a system to reach nearby infrastructure, it may be possible to offer alternate routes. In such a world, QoS and Traffic Engineering (TE) may be added to the raft of functions pushed into the routing. Even in the pure Ad Hoc environment, given MobileMAN includes a social element, where users are part of a trust and reputation system, it may be possible to build agreements about group priorities, so that different users and traffic are given some forms of priority. Furthermore, MANETs could be composed by heterogeneous devices with different hardware capabilities and resources.

This scenario calls for a *suite* of routing protocols to be included into MANET design, rather than a single, one-fit-all, routing protocol. User data will be transparently routed via the most suitable protocol, based on the user device capabilities, and the trade-off between the QoS the user wants to get and the cost she is willing to pay for. For example, routing trough cellular networks (such as UMTS or GPRS) is an expensive solutions that assures high coverage and prompt data delivery, while routing through opportunistic networks represents a virtually free-of-charge service, suitable for delay-tolerant data (see Section 3). Including different kinds of routing protocols in MANETs improves services' availability, and ultimately helps in making the MANET technology suitable for end users.

In previous deliverables we have already presented several routing protocols that can be included in such a suite (e.g., OLSR and HSLS). Now, we introduce Landmark Guided Forwarding (LGF). LGF is a novel approach towards ad hoc routing, which is able to exploit even imprecise information about node locations. Such a routing protocol can be used between MANET nodes with some localisation capabilities, e.g., GPS.

Whilst many routing protocols utilise either topologically driven route optimisation, or geographically driven route optimisation, we claim that a more efficient approach is to leverage benefits from each, creating a hybrid approach towards routing, optimised around various local and global parameters. Landmark Guided Forwarding achieves the following:

- Lower average routing state maintenance across the node set.
- Reduced the spread of routing updates.
- Reduced stale routing entries.
- Adaptive to dynamic Ad Hoc mobility.

Unlike topological routing protocols such as DSDV, DSR and AODV [17, 18, 27], *Landmark Guided Forwarding* requires that every node only maintains a small amount of topological and position information about neighbours within a localised area. Routing is achieved by using locally optimised algorithms, requiring lower network overhead. If the packet destination resides within the local area, it is routed using the shortest path algorithm. Otherwise, when the destination resides outside the local scope, it is routed towards a geographically determined optimal Landmark node. Unlike position based forwarding schemes suchas GPSR and Face routing [14, 16], LGF does not rely upon the establishment of planar graph, but leverages on the local hybrid routing information available, thereby increasing the resilience to inconsistent device position information and lowering the overall system vulnerability to position errors [21].

3.3.1. Assumptions

We make a few assumptions commonly used by position based forwarding protocols. We assume that every node knows its own geographic position. This is not an unreasonable assumption since it is feasible to gather position information from GPS or another positioning system. Since LGF does not require high precision, short range distance measurements from Bluetooth devices or via IEEE 802.11 based ranging systems such as the Intel Place Lab system [22] are suitable alternatives to a GPS based system. We also assume a distributed location service like the Grid Location Service [24] is available for a source node to retrieve the geographic position of a destination node.

3.3.2. Protocol Description

Ad Hoc networks rely on nodes in the network to relay packets between a source and a destination on behalf of their peers. As a packet flows between the source and destination, LGF calculates the locally optimal path to the destination and applies the shortest path to the destination if is within the local area. In cases where the destination is not within the local area, LGF employs locally optimal routing towards the node that is geographically closest to the destination. The protocol iterates progressively. Once the packet is forwarded, it will reveal a new set of neighbours and a local optimal route towards the destination. Using this technique not only effectively sidesteps the scalability constraints associated with global optimal routing as used by existing MANET protocols, but also allows routing to be more adaptive to the ever changing MANET topology. The approach taken by LGF only requires advertisement of topological and geographical information to a node's neighbours that are within a few hops. Thus it localises state dissemination and reduces the overall load on the network. This also allows localised state to converge much faster by adaptive updates that regulates the neighbourhood state update frequency based on the surrounding network connectivity. In essence, these properties allow LGF MANETs to be extended to a larger environment and be more adaptive to dynamic Ad Hoc mobility than other MANET protocols.

In this section, we present various algorithms that form Landmark Guided Forwarding. The protocol consists of various components, namely: restrictive hybrid route advertisement, adaptive route advertisement, link failure recovery, next hop selection, path exploration, dead-end detection and loop avoidance. We describe each of these in turn in the later sections.

Restrictive Hybrid Route Advertisement

In order to retain a balance between timeliness of routing decisions and the overhead of route advertisements, we propose a pro-active routing scheme based on a localised hybrid routing table. Using this approach, information about a node's geographical position and local topology is disseminated to a limited topological area. We define each node's neighbours to be within a topological area defined by the perimeter P in number of hops. For each neighbour node j within P, node i maintains its position, xj, yj, zj, and additional information as a routing entry REij , in the routing table RTi. A routing entry REij is given below:

REij = {j, NextHopij, HopCountij,{xj; yj; zj},{xj’,yj’, zj’}, SeqNumij}

where j is the destination and is a globally unique node identifier of all nodes within P, and the NextHop is the identifier of adjacent node that a packet should be forwarded to in order to reach the destination which is HopCount hops away. A sequence number SeqNum:ij is associated with each entry to ensure timeliness. The position and velocity of the destination j, are {xj ; yj ; zj},{xj' , yj' , zj'} respectively. These attributes are used by the forwarding algorithm to resolve a local optimal path when destination address dstp of a packet p is not in RTi, for each j; dstp is different from j.

In order to explain the restricted hybrid routing advertisement process with node mobility, we use an example. Figure 3.3 shows a small Ad hoc network scenario where node 3 moves from its central position to a new position in the top right of the network, all other nodes remain stationary. We demonstrate the scheme by comparing the routing tables and the topological view of the network from node 5's point of view.

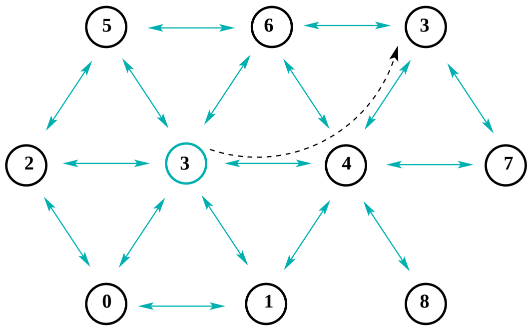


Figure 3.3. Mobility scenario in an Ad Hoc net-work

In this example, the restrictive hybrid advertisement does not propagate more than 2 hops from source and therefore nodes 7 and 8 are not included in node 5's routing table, Table 3.1, and its topological view of the network, as illustrated in Figure 3.4 before node 3 movement.

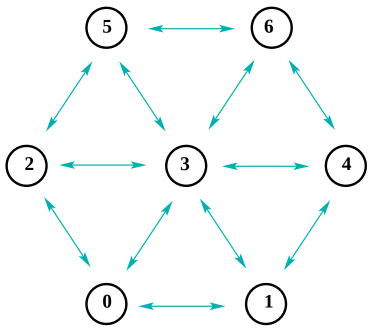


Figure 3.4. Node 5's topological view of the network before node 3 moves.

If we re-examine node 5's routing table, Table 3.2, and its topological view of the network, Figure 3.5, after the movement of node 3. Node 1 is now no longer routable using local optimal routing, nodes 3 and 4 are now only routable via node 6 and node 0 is only routable via node 2. In this scenario, node 3's movement causes the routing algorithm to make the following adjustments to the routing table of node 5, i.e. Table 3.1 is transformed to Table 3.2.

Remove entries for destinations which have a hop count greater than 2.

Update of location information.

Update of next hop and metric information.

Specifically it can be observed that the entry for node 1 has been removed from the routing table of node 5 in Table 3.2. The position of node 3 has been updated. The next hop and metric of nodes 3 and 4 have also been updated accordingly.

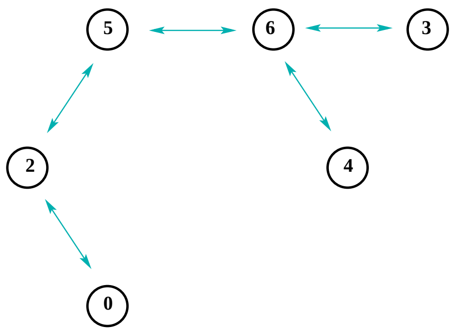


Figure 3.5. Topological view of node 5 after node 3 move away

Dst	Next Hop	Metric	x	y	z
0	2	2	300.00	2.00	0.00
1	3	2	450.00	2.00	0.00
2	2	1	225.00	132.00	0.00
3	3	1	375.00	132.00	0.00
4	3	2	525.00	132.00	0.00
5	5	0	300.00	262.00	0.00
6	6	1	450.00	262.00	0.00

Table 3.1. Node 5's routing table before node 3 moves.

Dst	Next Hop	Metric	x	y	z
0	2	2	300.00	2.00	0.00
1	3	2	450.00	2.00	0.00
2	2	1	225.00	132.00	0.00
3	3	1	375.00	132.00	0.00
4	3	2	525.00	132.00	0.00
5	5	0	300.00	262.00	0.00
6	6	1	450.00	262.00	0.00

Table 3.2. Routing table of node 5 after node 3 moves away

Next hop location algorithm

Our approach is to take advantage of the geographical position of those nodes that are within each node's topological scope as a basis for the forwarding algorithm. Each node i maintains the topological distance HopCountij and position xj ; yj ; zj for every other node j that is within its scope. The next hop is selected using the shortest path algorithm to each packet's destination d, where d matches one of the neighbours j. Otherwise, the next hop is determined by Landmark Guided Forwarding that selects the next hop by applying shortest path to a landmark node V. Where node V is geographically closer to the destination node D and topologically further away from node i. The term Landmark has been widely used to describe a physical point of reference for an Internet coordinate system [30]. In this paper, the Landmark is a temporary reference node amongst the collection of j, that acts as a virtual destination to assist in the routing of a packet towards its final destination. The exploration algorithm is progressive, as soon as the packet moves to the next hop, a new Landmark node amongst the new set of neighbours is determined and the packet progresses in the same manner until it arrives at a node with a topological path to the destination. However, in the case where no valid Landmark node is available for forward advancement, the path exploration algorithm rolls back and seeks an alternate path from the previous hop.

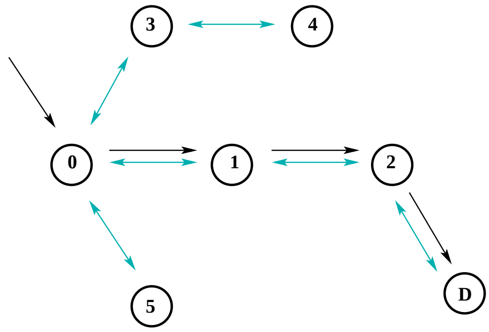


Figure 3.6. Next hop selection.

Figure 3.6 shows a subgraph that demonstrates our forwarding algorithm where the topological scope is limited to 2 hops. If we consider the packet arrives at node 0 destined for node 4, it can be forwarded to the destination via either node 1, node 3 or node 5, by applying the shortest path algorithm to the destination, the next hop is found to be node 3. In the case where a packet's destination is not within the coverage of the topological scope, the next hop is chosen by the shortest path algorithm to a landmark node. For this example in Figure 3.6, the next hop is node 1 since node 2 is found to be closer to the destination than node 4.

Path Exploration

In general, geodesic proximity to the destination does not assure a shorter topological path to the destination. Simply forwarding a packet towards its destination position without maintaining any forwarding path history does not provide any facility for preventing the packet being trapped by a localised loop or dropped due to a routing dead-end and subsequently backtracking. The approach adopted in our algorithm is to include a source path in the packet header and to also maintain soft forwarding state amongst all nodes traversed by a packet. By maintaining a source path in packet header it provides a trail of forwarding nodes such that in the event a dead-end is encountered, the packet can be back-tracked until it reaches a node with an alternative path to the destination. In addition, this also enables the algorithm to preserve its loop free property by not selecting a virtual landmark or next hop that is the source path. The purpose of maintaining soft-state within the network is to isolate and explore the network systematically. A node temporarily marks a link with the tuple {Packet Sequence Number, Next Hop, Soft State Expiry}, once it has forwarded a packet along that link. This enables the exploration algorithm to search all available paths and guarantee packet delivery where a path is available between a source and destination.

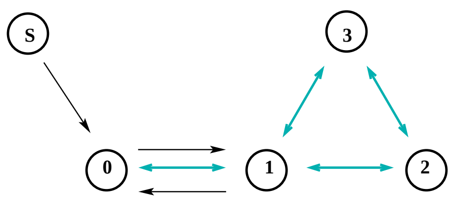


Figure 3.7. Dead end detection and roll back.

Figure 3.7 shows a subgraph that demonstrates how a dead end can be detected while a packet systematically explores a path to the destination. In this scenario, a packet from node S arrived at node 0. Assume the packet's destination is not reachable by any node in the figure. In addition, the destination is geographically closer to node 2 than node 3, the topological scope being 2 hops. We denote SP as a sequence of nodes in the source path. At node 0, where SP = (S), we determine the next Landmark node, according to our next hop selection algorithm, as node 2. The next hop node chosen to forward the packet towards node 2 is node 1 based on the shortest path algorithm.

When the packet arrives at node 1, SP = (S; 0), it becomes apparent that the only node that is 2 hops away from node 1 is S. However, since S is found in the source path SP, the algorithm considers S to be an invalid landmark. With no available landmark and the packet's destination not within the topological range, the path exploration detects that the packet is moving towards a dead-end and retracts the packet back to node 0. In this example, node 0 established soft-state when the packet was forwarded from node 0 to node 1 and likewise node S had established soft-state when the packet was forwarded from node S to node 0. Retracting back to node 0, the packet's source path SP is shortened to (S). At this point, the path exploration is aware that the link between node 0 and node 1 has already been visited. Since there is no forwarding path available, the packet is pulled back to node S. With no other link available at node S, the path exploration has exhausted all searches and drops the packet.

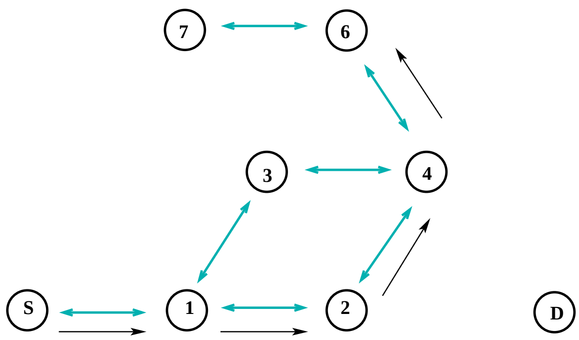


Figure 3.8. Loop avoidance.

Figure 3.8 shows a subgraph that demonstrates how a loop is avoided while a packet explores a path to its destination. In this scenario the topological scope is 2 hops and the source node is S. The destination D is not directly connected to any node in the subgraph. Based on our next hop selection algorithm, the packet at node S identi┐ es node 2 as its Landmark node. Following the shortest path algorithm to node 2, the packet is directed towards node 1. Subsequently, the packet is forwarded to node 2 with Landmark node 4.

The same process is repeated when the packet moves from node 2 to node 4 with node 3 as its respective Landmark node. When the packet arrived at node 4, it found SP = (S; 1; 2) with both node 1 and 7 at its' topological range, i.e. within 2 hops of node 4. With node 1 in its' source path, the algorithm provides only one option of forwarding towards node 6 with node 7 as the Landmark node. This effectively avoids the creation of a loop between 1-> 2 -> 4 -> 3.

Link Failure Recovery

When a node moves out of range of it's neighbours, established links are likely to break. Typically a broken link may be detected either by the link layer protocol timing out a connection, or it may be inferred at a higher level through the loss of a periodic broadcast signal which is expected within a predefined time. In our protocol, a node represents a broken link with infinity.

Figure 3.9 illustrates a mobility scenario in which node 3 moves out of range of nodes 0, 1 and 2. Node 1 is initially a neighbour of node 3, and records a route to node 3 with a metric of 1 as shown in Table 3.3. As node 3 moves out of range, the node detects the loss of a link, and updates it's table accordingly. Table 3.4 illustrates the change in routing metrics; the routes to both node 3 and node 2 which originally travelled via 3 are set to infinity. Node 1 subsequently broadcasts these routing entries to all it's single hop neighbours. Once the routing state has been synchronised in this manner, the node performs a periodic state maintenance process, removing or replacing the entries of1metrics with cheaper routes as illustrated in Table 3.5.

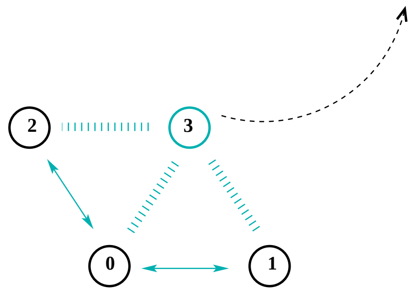


Figure 3.9. State Propagation and Maintenance.

Dst	Next Hop	Metric	x	y	z
0	0	1	300.00	2.00	0.00
1	1	0	450.00	2.00	0.00
2	3	2	225.00	132.00	0.00
3	3	1	375.00	132.00	0.00

Table 3.3. Routing table of node 1 before link broken

Dst	Next Hop	Metric	x	y	z
0	0	1	300.00	2.00	0.00
1	1	0	450.00	2.00	0.00
2	3	∞	225.00	132.00	0.00
3	3	∞	375.00	132.00	0.00

Table 3.4. Routing table of node 1 after link broken

Dst	Next Hop	Metric	x	y	z
0	0	1	300.00	2.00	0.00
1	1	0	450.00	2.00	0.00
2	0	2	225.00	132.00	0.00

Table 3.5. Routing table of node 1 after state maintenance

Adaptive Route Advertisement

One key feature of LGF is its ability to regulate the restrictive hybrid update frequency according to its interconnectivity with other adjacent nodes within its radio range. The essence of this feature is to associate update frequency with the furthest adjacent node. The greater the distance between node and its furthest adjacent node, the more frequently the node must send out its routing updates. This increase in the rate of state propagation enables the network to converge much faster when adapting to changes in the surrounding network connectivity.

3.3.3. Related Works

Many wireless Ad Hoc wireless routing protocols have been proposed in recent years. An early survey paper [26] categorised these protocols as table driven or source driven. In general, table driven protocols pro-actively gather topological routing information while source driven protocols reactively discover a route or routes to the destination as requested by the source. Pro-active routing protocols such as DSDV, Destination Sequenced Distance Vector [17], proactively exchange routing information between neighbouring nodes. The associated routing state and the network traffic overheads is O(n), where n is the number of nodes in the network, which does not scale well in large networks. Reactive routing protocols such as AODV [21] Ad Hoc on Demand Distance Vector, use flooding techniques to discover new routes and repair existing routes. As the amount of traffic in the network increases or the diameter of the network increases, the cost of flooding increases. With reactive routing protocols, the routing performance degrades under moderate mobility conditions [23,29].

An alternative approach to Ad Hoc routing is to take advantage of the physical location of nodes in the network and to do position based forwarding. An assumption made by protocols that take this approach is that every node knows its own geographical position. By limiting the exchange of positional information to be only between adjacent nodes, the state and network overheads are reduced to O(u), where u is the number of adjacent nodes. GPSR, Greedy Perimeter Stateless Routing [14], is a position based routing protocol that in general uses the geographically closest node to the destination as the next hop for the packet to be forwarded. However, in a local maximum scenario, this technique can prevent greedy forwarding from advancing towards the destination. To address a situation like this, GPSR uses a perimeter forwarding scheme that uses the well known right hand rule on its planarised graphs. Although GPSR scales well and is able to adapts to random topologies, it is vulnerable to position errors. A recent research article suggests that position inconsistencies in position based forwarding protocols could cause false greedy forwarding and misconstruction of the planar graph [21]. The results show various position inconsistencies do have significant negative effects on the performance of position based routing protocols. In LGF, position information is used together with a heuristic technique to explore the network systematically, our report [25] shows this approach is able to sustain significant position inconsistency without degrading routing performance.

LGF is similar in some of its features to existing routing protocols, such as ZRP [21] and Terminode [15]. In common with these two protocols, LGF uses a hierarchical framework that employs two different routing schemes. Each node pro-actively maintains connectivity with other nodes within its neighbourhood. A packet is routed using the shortest path algorithm when the destination is within this neighbourhood. In contrast, a packet destined for outside the local neighbourhood is routed using a more scalable routing protocol. ZRP uses reactive routing to determine the optimal path to destination whereas Terminode uses position based forwarding to forward the packet towards the direction of destination. In contrast to ZRP which uses a flooding technique to discover the destination, LGF progressively uses hybrid routing to explore the network systematically when delivering packets to the destination. In the case of the Terminode routing protocol, it uses greedy forwarding to forward packets, but it requires some static nodes to establish stable paths when greedy forwarding is not applicable. Conversely, LGF requires no static node, and its exploration algorithm is able to avoid looping and dead-ends even in the presence of high rates of mobility.

3.3.4. Simulation Scenario

The simulations have been carried out using the NS2 simulator [19], with each simulation lasting for 900 seconds. Each node uses the IEEE 802.11 MAC and the physical model models the radius of the radio range as being 250 meters. To fairly compare LGF with other analysis carried out in the literature, the simulation uses the random way point model to model node mobility. LGF evaluation with more realistic mobility models is ongoing. In all simulation scenarios, each node selects a random destination and moves at a speed uniformly distributed between 0 and maximum velocity. Upon reaching the destination, the node pauses for a configured period before it selects the next random destination and moves on. The traffic model uses constant bit rate UDP traffic flows, with 512 byte payloads. The start time for the different flows is uniformly distributed between 0 and180 seconds with each of the 30 traffic sources sending at the rate of 2 packets per second. In common with other protocol evaluations, [14, 17, 27, 18], we run several mobility patterns with different pause times at a constant maximum velocity. We use 5 different sets of mobility patterns generated with different pause times of 0,30,60,120,600 and 900 seconds where the maximum velocity is 15 m/s. This simulation is run in a geographic area of 1500x300 m2 with 50 nodes randomly placed. We also run simulations using different maximum velocities (1, 2.5,5,7.5,10,12,5,15 m/s) where the nodes move continually, with a pause time of 0 seconds. This second set of simulations is run in an area of 1500x 500 m2 with 100 nodes randomly placed in the area. We compare LGF with DSDV, AODV and GPSR using the different simulation scenarios we have just described and we compare the adaptability, performance and overheads of LGF with other MANET routing protocols. Each of the different MANET routing protocols has some settings specific to it, we detail these in Table 3.6, Table 3.7, Table 3.8 and Table 3.9.

Parameter	*Value*
Beaconing interval	3 s
Random variation of beaconing interval	0.5 %
Beacon expiration interval	13.5 s
Promiscuous mode	enable
Removal of neighbor from neighbor list when link broken	enable
Perimeter mode	enable

Table 3.6. GPSR specific parameters

Parameter	*Value*
Initial weight settling time	6 s
Periodic update interval	15 s
Number of missed periodic updates before declaring link broken	3
Settling time weight	7/8

Table 3.7. DSDV specific parameters

Parameter	*Value*
Lifetime of a route reply message	10 s
Time for which a route is considered active	10 s
Time before route request message is retired	6 s
Time which the broadcast id for a forwarded route request is kept	6 s
Number of route request retries	3
Maximum route request timeout	10 s
Local repair wait time	0.15 s

Table 3.8. AODV specific parameters

Parameter	*Value*
Tuning factor	1.5
Max expiry	15 s
Min expiry	1.2 s
Topological scope	3

Table 3.9. LGF specific parameters

3.3.5. Results

The results are divided into three subsections: performance with varying pause time, performance with varying velocity and path length.

Performance with varying pause time

Figure 3.10 evaluates the reliability of packet delivery of the different routing protocols; LGF, GPSR, AODV and DSDV. In general, DSDV, GPSR and AODV perform better as the pause time used in the random way model increases. In contrast, LGF is more robust at higher mobility levels, although the results indicate that its packet delivery ratio is relatively poor when compare to other protocols at low mobility. This is largely due to the way in which LGF handles link failures. GPSR, AODV and DSDV optimise the handling of link failure for stale connectivity. In contrast in LGF, we drop packets as soon as we see the link fail. It was a design decision to decrease the average packet delay at the cost of reducing the delivery ratio, we describe this in more detail below.

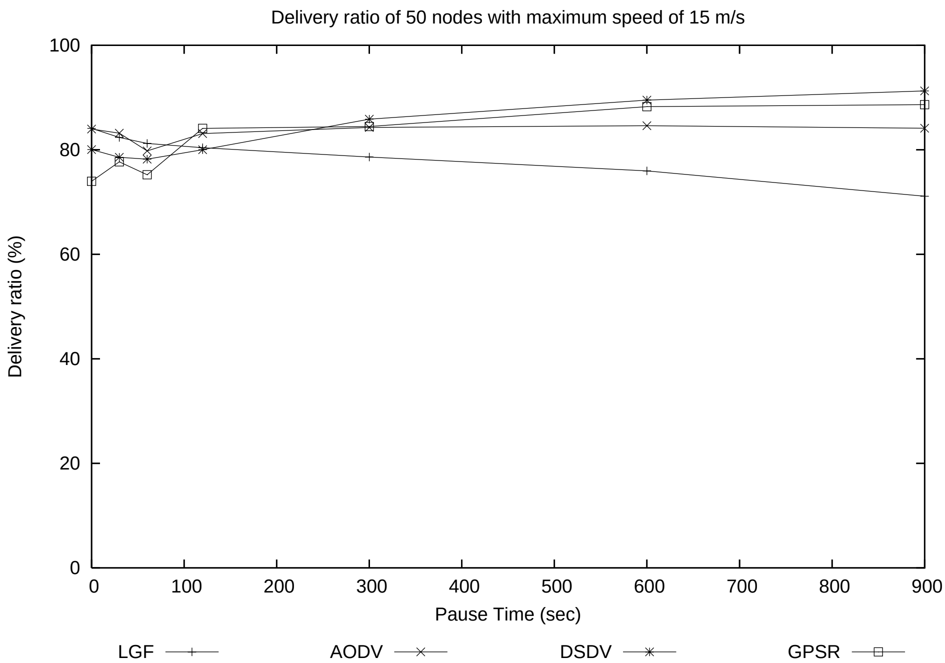


Figure 3.10. Comparison of the packet delivery ratio as a function of pause time

Both DSDV and AODV, upon notification of link retransmission failure, both protocols keep the packets in the buffer queue until the route becomes available again. This technique has not been published but it was found to be in the NS2 implementation. In the event of a link retransmission failure, GPSR applies the same technique used by DSR. It removes the routing entry of broken link before it en-queues the packet in the buffer for the routing protocol to forward the packet to a different next hop [14]. In LGF, the protocol drops the packet, updates the route entry, and propagates the broken link to other neighbouring nodes. Our results show that the link failure techniques used by GPSR, DSDV and AODV are opportunistic. The idea is to keep or redirect the packet when a link retransmission failure is encountered. Although this could increase the packet delivery ratio in some cases when connectivity is stable. However, in some scenarios such as where there is node mobility and the opportunity of direct or indirect re-delivery are not available, undelivered packets then linger for too long in the output buffer queue and can contribute to a higher average packet delay. Interestingly, our results show that the other protocols gain an advantage in the scenarios which use pause times of 300, 600 and 900 seconds. Current LGF design is unoptimised, we would expect to improve the performance of LGF in this respect.

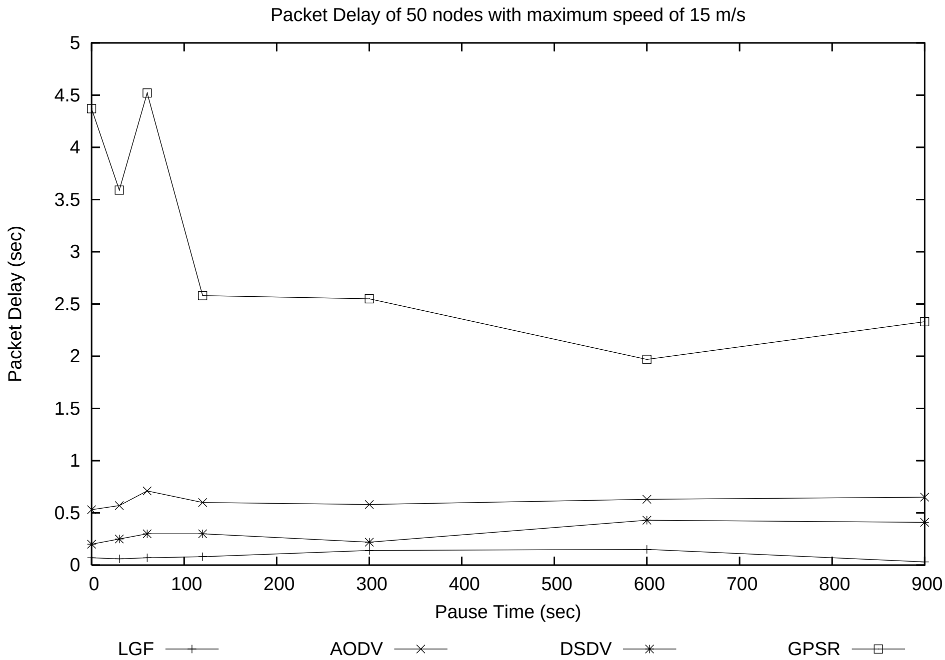


Figure 3.11. Comparison of average packet delays a function of pause time

This optimisation for increased packet delivery however does have side effects. From our observations, the average packet delay is increased as a results of this opportunistic delivery. In Figure 3.11, we show the effects on both AODV and DSDV are less significant as they only keep the undelivered packet for a short period of time. In contrast, GPSR retains the packet for much longer, this causes GPSR to have an increased delivery ratio, but this has the side effect of a higher average packet delay. Our results show LGF consistently has a lower latency than other routing protocols. LGF achieves this by not holding the packets in the event of link retransmission failure.

Figure 3.12 highlights the communications overhead of the different routing protocols. In LGF, the adaptive route update advertises more frequently in areas where link failure is more likely to occur while maintaining moderate updates in other areas where link failure is less likely to be encountered. Although, the advertisement is restricted to the local scope, LGF in general is sending out more frequent but restricted updates to its neighbours within its local scope. This explains why the overall communication overhead of LGF in this simulation is higher than DSDV. When compared with other protocols, LGF has lower communication overheads than reactive AODV but higher overheads than DSDV or GPSR. Despite its merit of having a low routing overhead, GPSR can encounter the effect of stale state when connectivity to its adjacent nodes changes more rapidly than its neighbours' periodic advertisements.

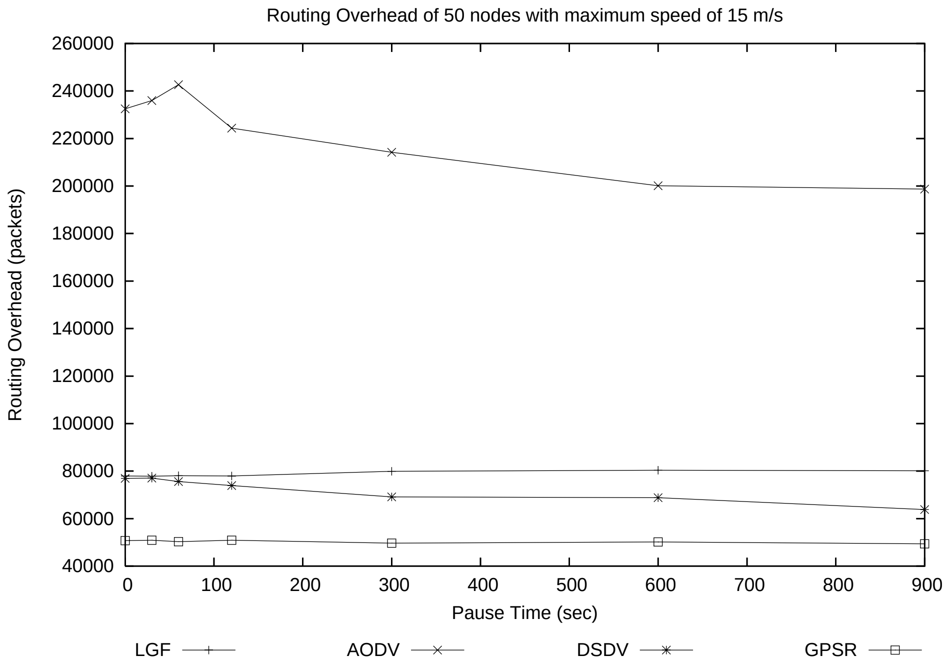


Figure 3.12. Comparison of routing overhead as a function of pause time

Performance with varying velocity

In this simulation, we tested performance of a system with 100 nodes over a wider area. Compared to previous simulations, the maximum distance between two nodes is larger, and therefore nodes are expected, on average, to take more hops between the source and destination. Additionally, the density of nodes in this simulation is 133 nodes per km2 as compared to the previous density of 111 nodes per km2. With more network overhead introduced as a result of the denser and larger system, it is further anticipated that contention and interference issues experienced in IEEE 802.11 networks could be more critical than previously measured. As a result, the channel capacity of the network is reduced [23] and consequently the average packet delay in general increases and the ratio of successful delivery decreases compared to previous simulations.

In comparison to other protocols, the results in Figure 3.13 however do indicate that LGF is relatively steady and robust with respect to the measured delivery rate over a variety of velocities. We can conclude from these results that LGF is more reliable and adaptive to unsettled, dynamic topologies than other protocols.

Our results in Figure 3.14 show that LGF performs consistently ell with respect to routing overhead over a variety of velocities. These results are similar to the previous simulation results, the high communication overheads associated with reactive AODV is a result of a higher number of route discoveries and local repairs AODV is performing. Comparing with earlier results where we used less nodes and a smaller physical area, the overheads we observed are more onerous than in the previous simulation. Our observations show that the overheads associated with LGF are lower than the other protocols as the number of nodes is doubled from 50 to 100. Because DSDV needs to maintain global state for all the nodes in the network, its overheads increase in proportion to the number of nodes in the network. In contrast, the restricted route update in LGF adapts well to the increased size of the network with the results confirming LGF's communication overheads scale better than the other protocols.

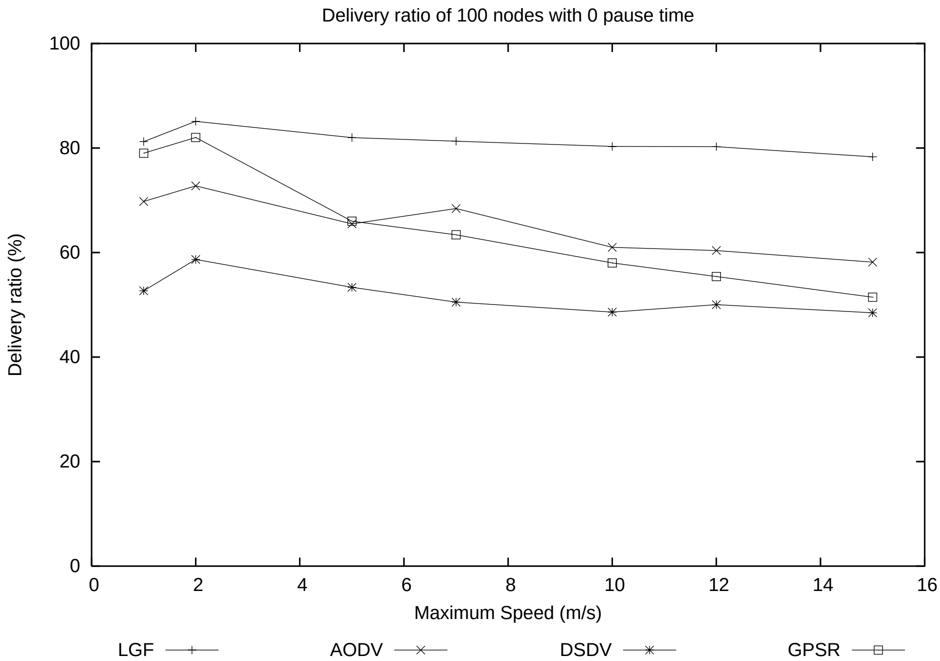


Figure 3.13. Comparison of packet delivery ratio as a function of maximum velocity

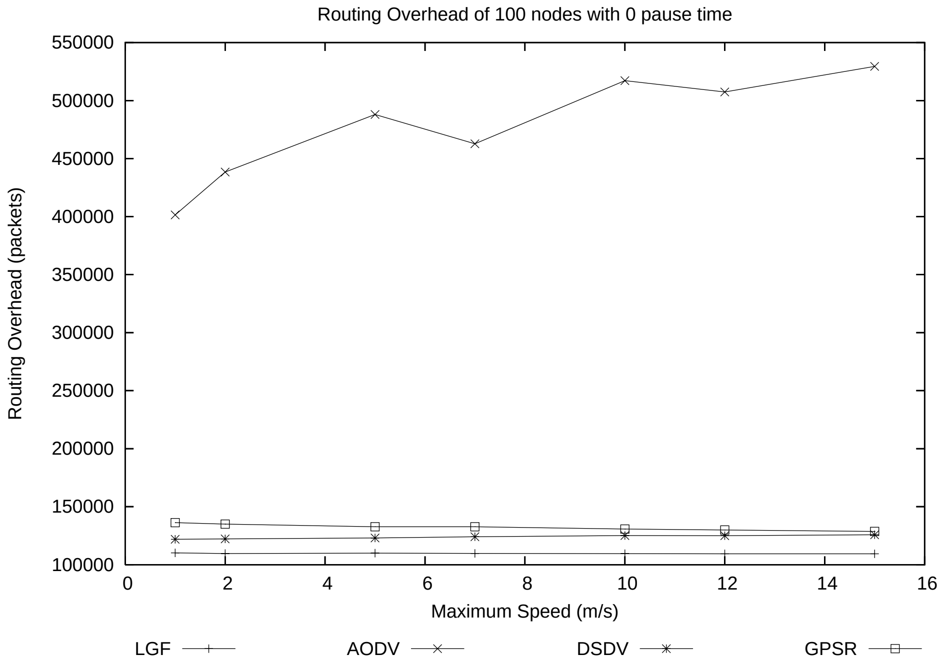


Figure 3.14. Comparison of the routing overheads as a function of maximum velocity

As shown in Figure 3.15, DSDV does not converge fast enough to cope with the changes in connectivity for the scenario where it uses a periodic update timer of 15 seconds, and when the network size has been increased. As a result of this, more undelivered packets are held in the queues in the network before they eventually expire and are dropped. Our results show the on demand path setup of AODV has a lower average packet delay than DSDV when simulating 100 nodes, this accounts for the performance advantage shown for the AODV local repair scheme in a dense network. If we consider the overall performance of all the protocol on packet delivery ratio, routing overheads and average packet delay, LGF provides a bette r overall balance performance than other protocols.

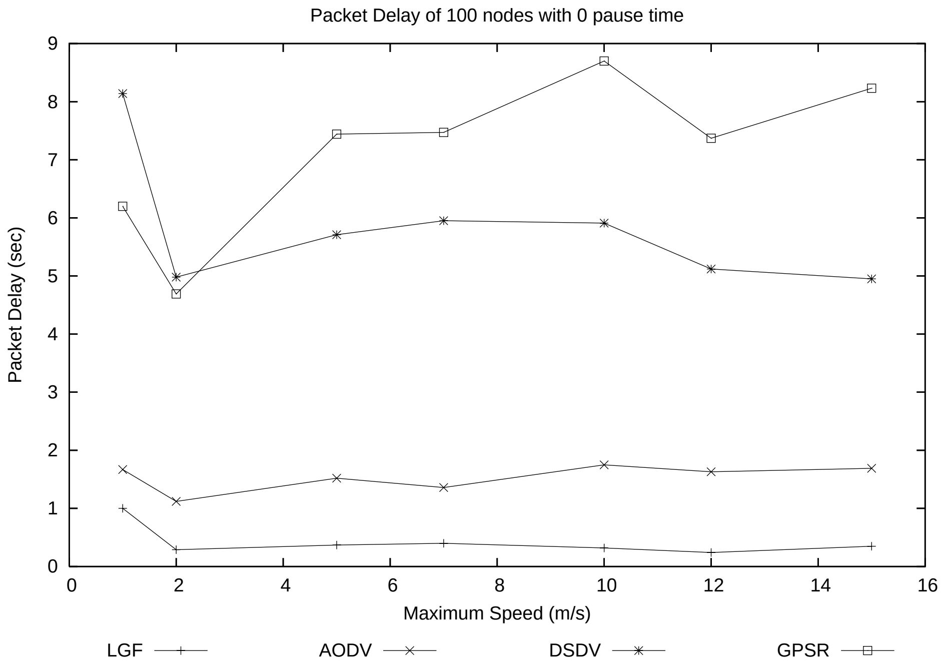


Figure 3.15. Comparison of average packet delay as a function of maximum velocity

Path length

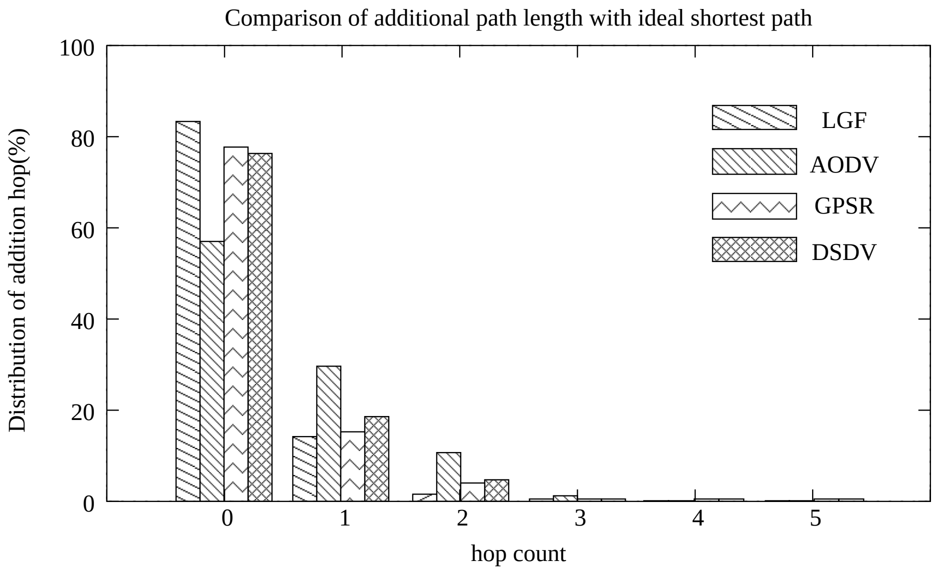


Figure 3.16 Comparison of the average path length for each of four protocols with ideal shortest path

Figure 3.16 compares the path length for successful delivered packets for each protocol against the ideal shortest path retrieved from the NS2 simulator. The ideal shortest path is the shortest possible path only constrained by the physical radio range. The evaluation was carried out with a random way point mobility model using a 0 seconds pause time with a maximum velocity of 15 m/s and 50 nodes placed randomly in area of 1500x300 m2. The results indicate that LGF on average achieves 83.52 % of optimal path length while GPSR obtains 78.98 % of optimal path length. Although theoretically DSDV is supposed to maintain an optimal path, the slow update interval does not prevent misleading stale state from being used by the packet delivery mechanism. This results in sub-optimal routing. DSDV only routes 77.37 % of its packets via the optimal path. Only 55.43 % percent of AODV's packets are routed by the optimal path. A contributing factor to this is AODV's local repair algorithm which is fixing broken paths without considering what the alternative optimal path between the source and destination is.

3.3.6. Concluding remarks and Future Works on LGF

In summary, we present a hybrid routing protocol, Landmark Guided Forwarding, which uses a restrictive hybrid advertisement at a rate regulated by its connectivity sensitive algorithm. LGF applies optimal routing when the destination is within its topological range, and systematically resolves a transient next hop through locally optimal resolution when an optimal route is unavailable.

We ran simulations with 50 nodes and 100 nodes, the results indicate the overheads of LGF scale better than other protocols when the number of nodes is double from 50 to 100. In our performance evaluation of varying pause time, it is apparent that route optimisations by AODV, DSDV and GPSR do improve the packet delivery ratio when rate of change of the topology is low, when using a mobility model where the pause time is greater than 120 and has a maximum velocity of 15 m/s. However the simulation results allow us to conclude that these optimisations could have the side effect of a higher average packet delay. The effect is more pronounced when simulated with 100 nodes, and a slightly wider network diameter. In contrast, LGF is able to maintain a steady, swift and reliable delivery even in the presence of a higher probability of unstable network connectivity. When comparing the path length with other protocols, LGF, surprisingly, has the highest score of the protocols under consideration.

In conclusion, local optimal routing sidesteps the constraint of maintaining a globally optimal path, as generally required by existing MANET protocols. This results in LGF being a relatively scalable and robust protocol with low overheads as compared to other MANET routing protocols, and yet, somewhat surprisingly, it retains relatively short routes nonetheless.

In future work, we wish to look into formal verification and setting up a testbed for LGF. In addition, we would like to investigate the use of a coordinate system with the Landmark Guided Forwarding protocol, to exploit the common goals of reducing routing overheads. Internet coordinate systems such as Lighthouses [28] and Virtual Landmarks [30] could be used to supplement the process of selecting the topologically closest node to the destination. LGF could additionally exploit the topological data from the coordinate systems to avoid routing errors when removing edges or nodes that violate the triangle inequality.

3.4. References

[1] R. Patra S. Jain, K. Fall. Routing in a delay tolerant network. In Proceedings of ACM SIGCOMM, 2004.

[2] E. Royer and C-K Toh. A review of current routing protocols for ad-hoc mobile wireless networks. IEEE Personal Communications Magazine, pages 46-55, 1999.

[3] M. Balazinska and P. Castro. Characterizing Mobility and Network Usage in a Corporate Wireless Local-Area Network. In 1st International Conference on Mobile Systems, Applications, and Services (MobiSys), San Francisco, CA, May 2003.

[4] M. McNett and G. M. Voelker. Access and mobility of wireless pda users. Technicalreport, Computer Science and Engineering, UC San Diego, 2004.

[5] T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. In Proceedings of ACM Mobicom, 2004.

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

[7] A. Jardosh, E. M. Belding-Royer, K. C. Almeroth, and S. Suri. Towards realistic mobility models for mobile ad hoc networks. MobiCom, San Diego, CA, 2003.

[8] M. Grossglauser and D. Tse. Mobility increases the capacity of ad hoc wireless networks. IEEE/ACM Transactions on Networking, 10(4):477-486, 2002.

[9] G. Sharma and R. R. Mazumdar. Scaling laws for capacity and delay in wireless ad hoc networks with random mobility. In Proceedings of ICC 2004. IEEE, 2004.

[10] J. Su, A. Chin, A. Popivanova, A. Goel, and E. de Lara. User mobility for opportunisticad-hoc networking. In Proceedings of WMCSA 2004, 2004.

[11] P. Juang, H. Oki, Y. Wang, M. Martonosi, L-S Peh, and D. Rubenstein. Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet. In Proceedings of the 10th International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS), pages 96-107, 2002.

[12] T. Small and Z. J. Haas. The shared wireless infostation model - a new ad hoc networking paradigm (or where there is a whale, there is a way). In Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pages 233-244,2003.

[13] P. Bremaud. Markov Chains, Gibbs Field, Monte Carlo Simulation and Queues. SpringerVerlag, 1999.

[14] B.Karp and H.T.Kung. GPSR: Greedy perimeter stateless routing for wireless networks. In Proceedings of 6th Annual Internation Conference on Mobile computing and Networking (MobiCom 2000), Boston, MA, USA, pages 243–254, Feb 2000.

[15] L. Blazevic, S. Giordano, and J.-Y. L. Boudec. Self organized terminode routing. Cluster Computing, 5(2):205–218, 2002.

[16] P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia. Routing with guaranteed delivery in ad hoc wireless networks. Wireless Networks, 7(6):609–616, 2001.

[17] C.Perkins and P.Bhagwat. Highly dynamic destination sequenced distance-vector routing(DSDV) for mobile computing. Computer Communication Review, 24(4):234–244, Oct 1994.

[18] D.Johnson, D.A.Maltz, and J.Broch. DSR: The Dynamic Source Routing Protocol For Multihop Wireless Ad Hoc Networks, Ad Hoc Networking. Addison-Wesley Longman Publishing Co. Inc., Boston, MA, 2001.

[19] N. group at ISI. Ns 2 home page. World Wide Web, http: //www.isi.edu/nsnam/ns/.

[20] Z. J. Haas and M. R. Pearlman. ZRP: a hybrid framework for routing in Ad Hoc networks. Book chapter in Ad Hoc Networks, Editor C. Perkins. Addison-Wesley Longman Publishing Co., Inc., 2001.

[21] Y. Kim, J.-J. Lee, and A. Helmy. Modeling and analyzing the impact of location inconsistencies on geographic routing in wireless networks. SIGMOBILE Mob. Comput. Commun. Rev., 8(1):48–60, 2004.

[22] A. LaMarca, Y. Chawathe, S. Conolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, J. Tabert, P. Powledge, G. Borriello, and B. Schilt. Place lab: Device positioning using radio bacons in the wild. Technical report, Intel Research, 2004.

[23] J. Li, C. Blake, D. S. J. D. Couto, H. I. Lee, and R. Morris. Capacity of ad hoc wireless networks. Mobile Computing and Networking, pages 61–69, 2001.

[24] J. Li, J. Jannotti, D. De Couto, D. Karger, and R. Morris. A scalable location service for geographic ad hoc routing. In Proceedings of the 6th ACM International Conference on Mobile Computing and Networking (MobiCom '00), pages 120–130, Aug. 2000.

[25] M. H. Lim, A. Greenhalgh, J. Chester┐eld, and J. Crowcroft. Hybrid routing: A pragmatic approach to mitigating position uncertainty in geo-routing. Technical report, University of Cambridge, 2005.

[26] E. M.Royer and C. K. Toh. A review of current routing protocols for ad hoc mobile wireless networks. IEEE Personal Communications Magazine, pages 46–55, April 1999.

[27] C. Perkins and E. M. Royer. Ad hoc on demand distance vector routing. In Proceedings of the 2nd IEEEWorkshop on Mobile Computing Systems and Applications, New Orleans, LA., pages 90–100, February 1997.

[28] M. Pias, J. Crowcroft, S. Wilbur, S. Bhatti, and T. Harris. Lighthouses for scalable distributed location. In Second International Workshop on Peer-to-Peer Systems (IPTPS '03), Feb 2003.

[29] C. E. S.Dass and E. M.Royer. Performance comparison of two on-demand routing protocols for ad hoc networks. IEEE Personal Communication Magazine, special issue on mobile Ad Hoc networks, 8(1):16–29, Feb 2001.

[30] L. Tang and M. Crovella. Virtual landmarks for the internet. In Proceedings of Internet Measurement Conference 143– 152, Miami Beach, FL., October 2003