Data Dissemination in Partially Cooperative Opportunistic Networks

A thesis submitted in partial fulfilment

of the requirement for the degree of Doctor of Philosophy

Abdolbast Greede

2013

Cardiff University
School of Computer Science & Informatics

Declaration

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed .... (candidate)

Date ....

Statement 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed .... (candidate)

Date ....

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed .... (candidate)

Date ....

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date ....

Abstract

Wireless communication between mobile users has become more popular than ever in the last decade, leading to increasing demand for network infrastructure. The growing popularity of smartphones among mobile users, leads an alternative infrastructure-less networking paradigm known as opportunistic networks. In opportunistic networks, mobile nodes such as smartphones use the mobility of devices in addition to wireless forwarding between intermediate nodes to facilitate communication without requiring a simultaneous path between source and destination. Without guaranteed connectivity, the strategy for data delivery is a key research challenge for such networks.

In this research, we present the design and evaluation of the Repository-based Data Dissemination (RDD) system, a communication system which does not rely on cooperation from mobile nodes but instead employs a small number of well-placed standalone fixed devices (named repositories) to facilitate data dissemination. To find the optimal location for their repositories, RDD employs knowledge of the mobility characteristics of mobile users. To evaluate RDD, a new mobility model "Human mobility model" has been designed, which was able to closely mimic the users' real mobility, and proven by conducting a series of experiments compared with real mobility traces. Using this model, the performance of the RDD is evaluated using custom simulation. In comparison with epidemic routing, the results show that RDD is able to drastically reduce resource consumption, expressed in terms of message redundancy, while preserving the performance in terms of data object delivery.

Acknowledgements

Above all, praise and thanks to Allah Almighty for providing me with patience, strength and persistence to complete this work.

This work would have not been possible without the continuous support, advice, and encouragement from my supervisor, Dr Stuart Allen. Throughout my research life during PhD, he has always been tireless in giving me invaluable comments and advice. His wealth of knowledge, sage advice and friendship made this research fruitful as well as enjoyable. Asides from work, he has also been exceptionally understanding and sympathetic with other problems I have encountered during this time. The support and encouragement provided by Dr Allen, has been invaluable on both academic and personal level, for which I am extremely grateful.

I am also very grateful to my second supervisor, Professor Roger Whitaker for his invaluable advice, support, and encouragement. I am also grateful to him for reading this thesis and providing invaluable feedback.

The support and advice provided by the staff of the School of Computer Science & Informatics, has been invaluable, especially Mrs. Helen William for her continuous support and her professionalism and Mrs Kirsty Hall, Mr. Robert Evans, and Dr. Rob Davies for their support. My thanks also go to my colleagues, Sultan Alyahya, Ehab ElGindy, Abdulhamid Elwaer, and Khaled Almuzaini for their friendship.

Last but not least, I must thank my family for their never-ending support, especially my mother Zinab, wife Khmila and my daughters Nora and Mayar.

Contents

List of Publications

Part of the work introduced in this thesis has been published in the following publications.

- Abdolbast Greede, Stuart M. Allen, and Roger M. Whitaker. A simple human mobility model for opportunistic networks.Post Graduate Network Symposium (PGNet), UK, 2008.
- Abdolbast Greede, Stuart M. Allen, and Roger M. Whitaker. RFP: Repository based forwarding protocol for opportunistic networks. IEEE Next Generation Mobile Applications, Services and Technologies International Conference(IEEE NGMAST), UK, 2009.
- Abdolbast Greede, Stuart M. Allen, and Roger M. Whitaker. RDD:Repository-based data dissemination protocol for opportunistic networks. International Conference on Selected Topics in Mobile and Wireless Networking (IEEE iCOST), Avignon, France, July 2012.

List of Figures

List of Tables

List of Algorithms

Abbreviations

1G First generation mobile phone system

2G Second generation mobile phone system

3G Third generation mobile phone system

4G Fourth generation mobile phone system

B Betweenness

CT Contact Time

DD Data Delivered

DR Data Redundancy

DS Data Sent

DTN Delay Tolerant Network

EPR EPidemic Routing

GSM Global System for Mobile Communications

H-RDD Hybrid Repository-based Data Dissemination

HMM Human mobility model

ICT Inter-Contact Time

ITU	International Telecommunication Union
MdNS	Mobility driven Network simulator
MN	Mobile nodes
OpNet	Opportunistic Network
RDD	Repository-based Data Dissemination
RWP	Random Way-Point mobility model
S-RDD	Scalable Repository-based Data Dissemination
SB	Social Betweenness

Mathematical Notation

Symbol	**Meaning /definition**
G	Graph that represent a map
V	Set of vertices within the graph G
E	Set of edges of G
H	Set of vertices that represent home locations
J	Set of vertices that represent junctions
$g(v)$	Degree of vertex v
M	Set of mobile nodes
$h(m)$	Home location for mobile node m
R	Set of repositories
$H(v)$	Maximum number of home locations can connect to v
$SpR(m, f)$	Spatial relationship between mobile node m to destination vertex f
P_{v_0,v_k}	Path between vertex v_0 to v_k
d	Data object
C	Subscription channels
$c(d)$	Channel assigned to a data object d
$m(d)$	Source of the data object d
$id(d)$	id of data object d

Symbol	**Meaning /definition**
$ct(d)$	Creation time for the data object d
$ttl(d)$	Time to life for data object d
$x(d)$	Expiration time for data objects d
$p(d)$	Priority of handling a data object d
$D_t(m)$	Set of all data objects stored at a mobile node m at time t
$D_t^{wt}(m)$	Set of data objects waiting to be send by mobile node m at time t
$D_t^{rc}(m)$	Set of data objects has been received by mobile node m at time t
$S(m)$	Channel subscription of mobile node m
$B(v)$	Betweenness value for vertex v
$SB(v)$	Social Betweenness value for vertex v
$\sigma_{v_s v_t}$	Number of shortest paths from vertex v_s to vertex v_t
$\sigma_{v_s v_t}(v)$	Number of shortest paths from vertex v_s to vertex v_t that pass through a vertex v
$C_T^s(m)$	Counter for the number of data objects that have been sent by a node m regardless of their destinations
$C_T^s(r)$	Counter for the number of data objects that have been sent by a repository r regardless of their destinations
$C_T^d(m)$	Total number of all data objects received by m where $c(d) \in S(m)$
$W(m)$	Willing to relay preference for mobile node m
L	Set of locations for R
$l(r)$	Location for repository r
$ExSB(v)$	Exclude social betweenness value for vertex v

Chapter 1

Introduction

This thesis considers the emerging communication paradigm of opportunistic networks. In particular, we propose a system of efficient push-based communication. In this chapter we introduce the area, motivate the study and summarise our contributions.

1.1 Background and Motivation

The use of wireless mobile devices continues to enjoy rapid growth as a direct result of recent technological advances in both the capability of devices and wireless communication technology. These developments have led to lower prices, higher data rates and a wider proliferation of wireless communication between mobile users. There are two fundamental approaches for enabling wireless communication between mobile devices. The first is infrastructure-based communication, where devices communicate with each other via a base station, which represents a key component of the network backbone (e.g. cellular network). This kind of network is limited to regions where such an infrastructure has been deployed and by the bandwidth provided by the network. The second approach is infrastructure-less wireless communication, where mobile devices communicate with each other either directly (when a pair of devices are within mutual transmission range) or indirectly (a multi-hop network). Most wireless mobile devices nowadays can support both types of wireless network, as can be seen in smartphones which are equipped with a GSM [35] (infrastructure), WiFi and Bluetooth (infrastructure-less) wireless networks.

The most widely used infrastructure wireless networks are cellular. When the first generation (1G) cellular network was commercially introduced in the USA in 1983 [35], voice communication was the main service provided. However, the great advances in wireless technologies have taken mobile devices beyond this classic use as a voice communication tool to encompass also data communication. In fact, the mobile users' demand for content is increasing particularly driven by streaming multimedia and the growth of user generated content, which has led to the cellular networks shifting toward data-oriented services. This direction is clearly seen in the third generation (3G) cellular network standard, which has a higher data rate and data-oriented services compared to the second generation (2G) standard, which has a low data rate and voice-oriented services [4, 35].

While the bandwidth provided by 3G to some extent solves the problem of the mobile users' data demands, the continuous advance in smartphones and the fast growing number of users for new social network applications, such as Facebook and Twitter, and an increase in user-generated content, i.e. more sharing between users, has led to increased demand for data. Thus is a need for a new cellular network standard that provides a better data rate. For this reason, in Octobar 2009, six technologies were submitted to International Telecommunications Union (ITU) seeking approval as the 4G communications standard [1]. In January 2012, two of the six technologies namely LTE-Advanced and WirelessMAN-Advanced (Known as WiMAX) were approved as an official platforms for 4G. 4G represents a big shift in data rate, which can peak at 1Gb/s compared to the 10Mb/s of 3G. While the 4G standard could solve the data rate problem, it still requires the mobile user to be attached to the network infrastructure to enjoy the benefits of the new standard. In addition to this infrastructure requirement, cellular network providers usually apply expensive charges for the use of their bandwidth. Although the cost problem could be solved at some stage, the infrastructure requirement is still hard to guarantee.

There are three scenarios in which the infrastructure network could be unavailable for mobile users. First, the infrastructure may simply not exist, as seen in many developing countries or in some rural areas of the developed world. In the second scenario, the infrastructure has been temporarily disabled, as can happen during any natural disaster such as an earthquake, tsunami etc. Thirdly, mobile users may find themselves not capable of reaching the network infrastructure, even it exists, as seen in the so-called 'Arab spring' countries in 2011 (Libya, Egypt, and Tunisia), where their dictators shut down the network operators, leaving the only options for users of infrastructure-less networks. Because of the extensive use of the internet by smartphone users, other problems start to appear, such as mobile phone providers running out of capacity to handle all of the data request, this can be particularly problematic in a scenario such as the Olympic Games, where hundred of thousands people come from abroad with their smartphones and tablets to temporarily join the local networks providers.

In classic infrastructure-less wireless networks, mobile devices communicate with each other either directly or indirectly. In a multi-hop network, every device can behave as a router, forwarding packets on behalf of other devices in the network. This type of network is known as an mobile ad hoc network [71]. Although there are a few implementations of ad hoc networks in the military sector, the end-to-end connectivity requirements make it harder for successful implementations in civilian sector.

Delay tolerant networking (DTN) is another infrastructure-less network paradigm which is capable of providing communication, that relaxes the need for end-to-end connectivity [27]. This is facilitated by information being stored at nodes and subsequently shared with peers when an opportunity to forward arises. When DTN was initially introduced, the popularity and advances of mobile devices was not at the stage they are today. Therefore, DTN was not specifically designed for mobile phones but in a generic form for heterogeneous wireless mobile devices that move from one place to another and employ that movement to deliver data.

In the last 5 years, a new infrastructure-less network paradigm has emerged, known as Opportunistic Networks (OpNets) [64]. In an OpNet, the wireless mobile devices are clearly defined as homogeneous mobile devices carried by individual people. Therefore, if we only distinguish OpNets from DTN by mobile devices and their carrier, OpNets could be considered as a subset of delay tolerant networks. However, there is another distinguishing characteristic between OpNets and DTN. In the DTN, the disconnection point is usually known in advance and sometimes it is even known how long this disconnection will last, while in OpNets these occur as a consequence of an individuals movement and they do not exhibit any clearly defined patterns, which is why OpNets are called opportunistic. Hence many researchers see OpNet as a general form of DTN.

A recent ITU report has estimated the number of mobile phone users by 6.8 billion subscriptions at the 2013 [81]. The fact that the population of this planet is only 7.1 billion, means that there is nearly a mobile phone for each human on the planet. The number of smartphone users is expected to continue to grow very rapidly within the coming years and on the other hand, the number of non smartphone users are expected to shrink. Smartphones are often equipped with short-range wireless connectivity, making OpNet a good solution for sharing data between smartphone users'. The OpNet enables users to share data between them without need to worry about the cost and doesn't need the end-to-end connectivity required by classic mobile ad hoc networks. However, the design of efficient data delivery protocols for OpNets is a complicated task due to the absence of knowledge about the topological evolution of mobile users. Although in the last few years, much research has been conducted in designing the data delivery protocol, the implication of the human aspect on the networks has not got enough attention. In this research, we consider more about the implications of human aspect.

1.2 Research problem

The focus of this work is information dissemination in OpNets rather than point-to-point communication, where communication is based on information push not pull. That is, the system should aim to forward relevant content to groups of interested users without explicit prior request for individual content. Sharing locally-relevant user-generated content is an example of a potential application where such push-based communication is more appropriate. This is because other than the user who generated the content, other users do not have any prior knowledge of what content has been generated and existing in the network. The main problem considered in this research is how to design an efficient and realistic data dissemination mechanism for OpNets. While many approaches have been introduced for data dissemination in OpNets, most of these were unable to achieve the practicality and efficiency required to use in real life. In this research, we propose an approach which aims to provide both efficiency and practicality for data dissemination. This includes a dissemination protocol which does not rely on high levels of cooperation, instead utilising a small number of well placed, standalone devices. To achieve accurate evaluation of the proposed system, a suitable evaluation tool is needed. Evaluation of such a protocol is usually carried out by simulation, using a mobility model to mimic human movement. Since none of the existing mobility models sufficiently mimic key aspects of human movement, a new mobility model is needed that is capable of mimicking human movement fairly accurately. This is another problem to be addressed here to enable the accurate evaluation.

1.3 Hypothesis

Due to the nature of OpNets components, i.e. mobile users carrying devices, two main assumptions should be taken into account in any proposed approach. The first is that since devices are carried by mobile users, their mobility patterns are identical to their carriers, which means the structure and regularity of movement can potentially be exploited. The second is that in practical use, there is limited incentive for users or devices to cooperate in forwarding traffic. This may be due to selfish behaviour in mobile users [39, 55] as a result of security awareness or the limited resources of their devices. Such selfish behaviour deters mobile users from carrying data on behalf of others. These two issues should be taken into account in designing an efficient and realistic data dissemination system for OpNets. Other important features should be taken into account are users privacy and system practicality. In contrast to existing schemes, we aim for a minimal and simple protocol, which avoids the need to share personal information to facilitate data dissemination. Based on the above, our hypothesis is as follows.

With appropriate protocols, characteristics of human mobility patterns can be used to place standalone fixed wireless devices to work as relay nodes for data dissemination in practical opportunistic networks resulting in:

- *Decreased reliance on mobile nodes cooperation.*
- *Reduced network overhead.*

1.4 Research Objectives

This research has the following objectives:

1. Develop an efficient data dissemination system for OpNets that;
    - Is practical to implement;
    - Does not rely on high levels of mobile nodes cooperation.
2. Validate the proposed system in a simulation evaluation tool that provides a mobility model mimicking the relevant aspects of human movement.
3. Investigate techniques to place relays in appropriate locations.

1.5 Research Contributions

This work proposes and evaluates Repository-based Data Dissemination (RDD), a system that exploits the user's mobility pattern characteristics and content preferences to drive the dissemination process. RDD introduces the use of standalone fixed wireless devices, referred to as repositories, to work as dedicated relays. These repositories are standalone fixed wireless devices which lack connection to any form of infrastructure networks (such as the internet), so can be placed and removed easily from any chosen location. This thesis takes a holistic approach, considering not only the development of the protocol, but also providing contributions related to protocol deployment and simulated evaluation. While this research does not builds a system prototype, the requirements for deployments of a real system were clearly defined. The location at which repositories are placed is clearly an important factor in the performance of RDD. A location may be considered suitable if it has a high rate of contact with mobile devices. Because mobile devices are carried by humans, the study of human movement patterns is conducted using simulations and real trace environments to optimise the locations for such repositories.

RDD can be considered the first of its kind to use a standalone wireless device. To evaluate the system using realistic human mobility characteristics, a novel mobility model has been designed. This model is able to generate movements which exhibit similar characteristics to human mobility combining the best features of existing approaches. To evaluate its realism, it has been validated against real traces comparing some key characteristics. RDD performance has been evaluated across a range of metrics using the epidemic routing protocol (EPR) as a comparison. Even with a single, well-placed repository, RDD is able to significantly reduce resource consumption. The results clearly indicate that RDD is a promisi ng method for data dissemination in OpNets, especially in real life scenarios where other approaches fail due to non-cooperative nature of mobile devices to carry messages.

1.6 Thesis Organization

The remainder of this thesis is organized as follows. We start with a review of the related literature in Chapter 2. We begin this Chapter with understanding the opportunistic networking concept. Then we discuss data delivery in opportunistic networks and we highlight the problems associated with current approaches. At the end of this chapter, a set of requirements are presented. Chapter 3 presents the evaluation tools used in this thesis. We start with presenting our custom simulator, then we review the available mobility models and present the main point of motivating the design of a human mobility model. At the end of the chapter we present the human mobility design and its validation. Chapter 4 presents the data dissemination system and its design. Chapter 5 demonstrates thorough evaluation of the system, using the proposed mobility model, and discusses the tradeoffs in terms of data delivery and overheads. Chapter 6 introduces RDD extensions which are Hybrid RDD and Scalable RDD. Chapter 7 summarizes the contributions of this work, outlining possible future research directions.

Chapter 2

Background and Literature Review

In recent years, OpNets have gained some attention which has contributed positively toward applying such networks in real life, but there remain several different research issues that need to be addressed before their mainstream adoption is practical. One of the main issues needs to be address is the data delivery. At the start of this chapter, we discuss infrastructure-less wireless networks in general and then focus on OpNets and their applications. After that, the data delivery problem and related protocols are discussed, and finally we address the limitations of the current approaches and what is needed to overcome these limitations.

2.1 Mobile Wireless Communication

One of the significant features of our life today is the great use of mobile devices by people. Wireless communication between mobile device users has become more popular than ever. The wireless radio connectivity available today goes back to in 1886 when electromagnetic waves were first discovered by Heinrich Hertz [14]. The electromagnetic spectrum is usually classified into different classes arranged according to frequency, which are: Gamma, X-ray, Ultraviolet, Visible, Infrared, Microwave, and Radio waves. The radio wave spectrum is made up of a range of radio frequencies bands, from extremely high frequencies (300 GHz) down to extremely low frequencies (3 Hz).

Regulators divide the spectrum into two types of bands, licensed and unlicensed. Licensed bands, require that individual companies pay a licensing fee for the exclusive right to transmit on assigned channels within that band in a given geographic area, while unlicensed bands don't require any permission, so long as products and users comply with the rules associated with that unlicensed band (for example, maximum transmission power). The 2.4 GHz Industrial Scientific and Medical (ISM) band is allocated in most countries for use by anyone, without a license. The Bluetooth network is an example of popular networks using this 2.4 GHz band.

Two approaches have been adopted for enabling wireless communication between devices, which are infrastructure-based networks and infrastructure-less wireless networks. In this research, we focus only on the infrastructure-less networks.
In infrastructure-less wireless networks, mobile devices communicate without needing a central infrastructure to be deployed. In this case, mobile devices communicate with each other either directly (device to device) or indirectly (a multi-hop network). In a multi-hop network every device can behave as a router, forwarding packets on behalf of other devices in the network. This type of network is known as a mobile ad hoc network, which is a typical example of an infrastructure-less network. Mobile ad hoc networks will be discuss in detail in the next section.

The type of mobile wireless device envisaged in this research is typically a mobile phone or tablet with short range wireless technology. This technology may be Bluetooth, ZigBee, WiFi, or any other technology operating in unlicensed bands [49]. Table 2.1 shows the main characteristics of these three main popular technologies, and we note it shows that they are all broadly similar for this purpose. Hence in this research, we consider a general model of transmission. There are many networking paradigms for infrastructure-less communications. In the following section we will discuss infrastructure-less networks in detail.

Technology	Bluetooth	ZigBee	WiFi
IEEE spec	802.51.1	802.15.4	802.11a/b/g
Frequency	2.4 GHz	868/915 MHz, 2.4 GHz	2.4 GHz
Max signal rate	1Mb/s	250 Kb/s	54 Mb/s
Range	10 m	10-100m	100m

Table 2.1: Bluetooth, ZigBee, and WiFi Characteristics

2.2 Mobile Ad Hoc Networks

Mobile ad hoc networks [71] are a typical infrastructure-less paradigm, and have attracted research attention for the last two decades primarily motivated by military applications, such as soldiers communicating in hostile area. An ad hoc network is defined as a collection of wireless mobile nodes dynamically forming a temporary network without the use of existing network infrastructure or centralized administration. Mobile devices are the only backbone of such networks. Figure 2.1 shows an example of group of soldiers on the battlefield who form an ad hoc network.

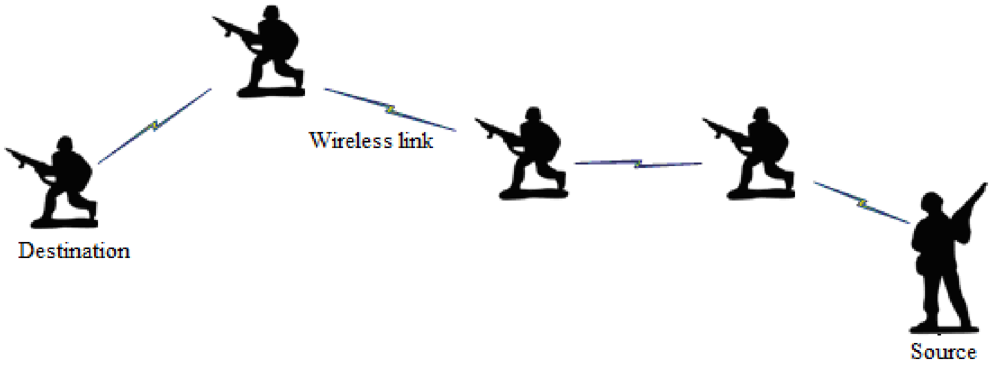


Figure 2.1: An example of an mobile ad hoc network

The IETF MANET (Internet Engineering Task Force Mobile Ad hoc Network) group defines mobile ad hoc networks as an autonomous system of mobile routers (and associated hosts) connected by wireless links - the union of which forms an arbitrary graph.

The routers are free to move randomly and organize themselves arbitrarily; thus, the network's wireless topology may change rapidly and unpredictably. Such a network may operate in a standalone fashion, or may be connected to the larger Internet.

2.2.1 Characteristics of mobile ad hoc networks

RFC 2501 [22], published by the IETF MANET group, has highlighted many of the important characteristics of mobile ad hoc networks, which are as follows:

1. **Mobility**: Mobile ad hoc networks usually have a dynamic topology due to the fact that nodes change their physical location by moving around. This means that a valid path at a given instant might not work at all a few moments later.
2. **Limited hardware**: Mobile ad hoc nodes usually have severely limited CPU capacity, storage, battery power and bandwidth. This mean that the power usage must be limited, which results in limited transmitter range.
3. **Unidirectional link**: In a wired network, transmission between two devices works equally well in both directions (bidirectional) but in a wireless environment this is not the case (e.g. two nodes may have different strengths in their transmitters).
4. **End-to-end connectivity network**: In an ad hoc network, devices need to be connected at the same time to the same network to be able to communicate directly or indirectly.

2.2.2 Mobile ad hoc network applications

Because there is no need for infrastructure to be in place, ad hoc networks are in a good position to be used for military applications, smart home applications and in disaster environments. However, mobile ad hoc networks require a valid path between the source and the destination to be maintained to support communication. Due to node mobility, these connections often hard to maintain. Routing in mobile ad hoc networks thus becomes a challenge; therefore many routing protocols have been introduced to handle this requirement.

Routing protocols in general can be divided into two main categories: proactive and reactive routing. Proactive routing protocols attempt to maintain consistent, up-to-date routing information from each node to every other node in the network. Usually these protocols involve each node sending out periodic messages, so that changes can be detected. Examples of proactive routing protocols include Dynamic Destination-Sequenced Distance-Vector Routing Protocol (DSDV) [65]. On the other hand reactive protocols only create routes when required. When a source wants to send to a destination, it uses the appropriate route discovery mechanisms to find a path to the destination. This route remains valid whilst the destination is reachable, or until the route is no longer needed. Examples of reactive routing protocols include Dynamic Source Routing (DSR) [44].

Because the mobile ad hoc network requires end-to-end connectivity, there are few commercial implementations and usage of this networks. However many of the beneficial properties of mobile ad hoc networks could be applied to other situations, e.g. direct communication when devices are in proximity and their mobility handling abilities.

2.3 Delay tolerant networks

To enable communication between mobile devices in environments where end-to-end connectivity is not possible, the Delay Tolerant Network (DTN) paradigm has been introduced. DTNs are a model of wireless infrastructure-less network that seeks to provide communication between devices without end-to-end connectivity required by classic TCP networks or mobile ad hoc networks [90]. The DTN has been a research topic for more than 10 years. It originally started when NASA was working on a way to expand Internet-like connectivity to situations such as communications with spacecraft on deep space missions [27]. The distances involved are huge, which means that the large round-trip delays make conventional Internet protocols such as TCP unusable. For example, 35 years after leaving Earth, the spacecraft Voyager 1 is now more than 18 billion kilometres away from the Earth in March 2013.

A typical DTN architecture consists of a network of independent networks, each characterized by Internet-like connectivity within, but having only occasional communication opportunities between them, sometimes scheduled, some completely random. Independent networks are located apart from each other, each internet relying on a protocol stack that best suits its particular infrastructure, communication means, and technologies. These independent networks form so-called “DTN regions”, and a system of DTN gateways is in charge of providing interconnection between them [64]. Hence, there are two general characteristics in a typical DTN system, which are: the points of possible disconnections are known and isolated at gateways, and each DTN could have a different network infrastructure, i.e. it is a heterogeneous network.

The DTN architecture assumes that nodes operating in a DTN environment do not necessarily guarantee end-to-end connectivity; therefore the communication paradigm can be described as store, carry and forward. In a DTN, data forwarded to a node is stored until another node is encountered that appears to provide a better chance of delivering the data to its intended destination.

While the DTN is the most widely used terminology to describe such networks, in the literature, these networks are found under different terminologies such as sparse mobile ad hoc networks [89] or intermittently connected networks [54], [78] .

2.3.1 Characteristics of DTN

While DTN shares some characteristics with mobile ad hoc networking, it differs from mobile ad hoc networks in two characteristics. The first is the limited resources of mobile devices is not always the case in DTN because sometimes the carrier is a vehicle or aircraft. The second main difference is the end-to-end connectivity requirement for mobile ad hoc networks is not valid for DTN. The last different is that DTN usually composed of heterogeneous nodes and protocol in sub networks.

2.3.2 DTN Applications

There are many examples of DTN applications, here we briefly discuss some examples to demonstrate the range of benefits. DakNet [2] is a wireless network that provides asynchronous communication to rural areas. This is facilitated by equipping motorcycles and buses to carry user's email and web search messages from remote rural residents. DakNet has been successfully deployed in remote parts of both India and Cambodia, achieving its goal of providing asynchronous networking service at very low cost.

In the most north part of Sweden, the Saami people live where herding reindeer is a source of employment (for 10 per cent) and of major cultural significance. This area served by the Saami Network Connectivity (SNC) project [6] aimed at providing asynchronous network connectivity. The SCN project goals are providing email, file transfer, cached web services, and herd tracking telemetry (which requires confidential reporting of herd movements to herders). The basic design of SNC involves setting each remote area as its own network region. At the edge of the network region a gateway. The data bundles are then relayed between gateways using DTN routing through a series of fixed and mobile relay caches.

Another example is a wildlife tracking application called ZebraNet [66] aimed at monitoring wild species. In ZebraNet, the network is constituted of small wireless devices that are attached to zebras, which log movement patterns of the zebras. When two zebras get into proximity, the wireless devices exchange collected data for potential data delivery back to a base station. The base stations are mounted on vehicles which move around sporadically. Another similar system has been proposed in [76], where the network is made up of a set of wireless devices attached to whales and a set of fixed nodes (infostations) act as collecting nodes. As we have discussed in this section the nature of the wireless device in DTNs is not clearly defined, which limits the nature of information that can be used to drive dissemination.

2.4 Opportunistic network

OpNets share the same fundamental concept with DTNs in that neither require the sender and recipient to be connected at the same time. This is facilitated by information being stored at nodes and subsequently shared with peers when the opportunity to forward arises. The simple form of opportunistic networking can be described as follows. If a node has information to be sent, this node stores the information until it encounters another node, when it forwards this information to it. If the recipient node is not the target node, the recipient node stores this information to be forwarded when it encounters other nodes. This forwarding process between encountering nodes is continuous until this information is delivered to its destination.

The OpNet uses a store, carry, and forward communication paradigm. At first a node stores data to be sent and after that it uses its mobility to carry the data and when another node is encountered, it forwards the data. Figure 2.2 shows an example of OpNets between people, highlighting the use of wireless communication in addition to user mobility for data delivery.

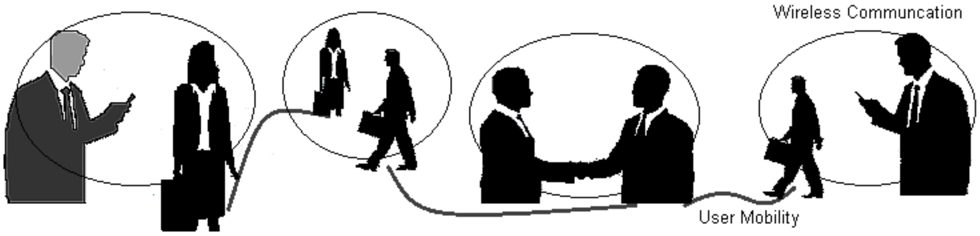


Figure 2.2: An example for OpNet

There is no commonly agreed terminology in the literature, or a clear separation of concepts for opportunistic and delay tolerant networks. The terms OpNet and DTN are often used interchangeably. In this thesis, we adopt the following definition:

> "An OpNet is an infrastructure-less wireless network that allows human-oriented mobile wireless devices, such smartphones, to communicate with each other. If devices are in mutual range, the communication is facilitated by direct wireless links. If the devices are not in mutual range, the communication is facilitated by human mobility and wireless links without the need for end-to-end connectivity between them."

2.4.1 Characteristics of OpNets

While an OpNet shares some concepts of other wireless infrastructure-less networks, it has some special characteristics which are as follows:

1. The wireless devices comprising the network are mobile devices carried by humans.
2. The mobility patterns for the devices are the same as the mobility patterns for their carriers which are independent and uncontrollable movement patterns.
3. The devices comprising the network have limited resources of battery, storage, and networking.

4. The communication between devices is through the network using only infrastructure-less wireless networks.

5. All wireless devices comprising the network have a short range wireless capability with the same radio technology, i.e. it is a homogeneous network.

6. Wireless devices are capable of communicating each other directly if they are in mutual range.

7. The communications between devices that are not in mutual range does not require end-to-end connectivity. This is facilitated by using the store, carry and forward communication paradigm explained above.

As we can see from these characteristics, there are some differences between DTNs and OpNets. In DTNs the wireless device is not clearly defined so it could be a bus or airplane, a spaceship or any other wireless device, whilst in OpNets it is clearly defined as a mobile device carried by humans. In DTNs, the place of possible disconnection is usually known in advance, but OpNets are more unpredictable. Furthermore, in DTNs the duration of disconnection is usually known while OpNets are designed assuming the duration is more unpredictable. Therefore, if OpNets were distinguished from DTNs by the nature of the mobile device carrier, OpNets could be classified as a subset of DTNs. However, if they are characterised by mobility pattern of mobile device, DTNs can be seen as a subset of OpNets, because the disconnection point and time in a DTN is often known in advance, while in the OpNet the mobile devices are carried hence their mobility is usually hard to predict. Because of the use of mobile device in OpNets, some researchers employ the term Pocket Switched Networks (PSN) as seen in [40],[62],[16].

2.4.2 OpNet applications

In view of the fact that the OpNet doesn't needs end to end connectivity required by mobile ad hoc network and the main components of OpNet are the mobile devices which are widely used among people today, OpNets look promising for the civilian sector. The main drawback of opportunistic networking techniques comes at the price of a lack of guaranteed delivery and additional delay, since messages are often buffered in the network waiting for a path towards the destination to be available. However, many applications could tolerate this. A further significant issue is the need for mobile nodes to cooperate to achieve data delivery. This requires devices to act altruistically for mutual benefit rather than selfishly to protect their own resources.

One of the interesting trials of building an opportunistic networking platform was started in 2006 through an European Commission project called Haggle [72]. Haggle was designed as network architecture to enable communication in the absence of end-to-end communication infrastructure. Although the Haggle project ended in 2010 it is still an important example of defining and building a complete opportunistic network platform.

2.4.2.1 Haggle Project

The main goal of the Haggle project was to provide a more human-oriented way of communicating. As highlighted above, some researchers describe OpNets as Pocket Switched Networks (PSN) which is the case in the Haggle project. PSN aim to allow applications to take advantage of all available types of wireless data communications without having to specifically code for each circumstance. PSN also aim to allow networking endpoints to be specified by user-level naming schemes rather than node specific network addresses. Because of limited resources available to mobile devices, PSN aim to use these resources efficiently. Members of the Haggle project have presented a set of interrelated principles which were used in their network design [72]. These principles are described in the following list:-

1. Forward using application layer information. Applications such as user name instead of address.
2. Support asynchronous operation.
3. Empower intermediate nodes which means intermediate nodes can keep a cache of the data transferred in case its own user later.
4. Message switching which means that intermediate nodes are provided with the whole message so they can act as destinations as well as forwarding points for any given message.
5. All user data kept network visible All user data should be made visible to Haggle at all times and these data must be marked with metadata about its user-level properties, such as access authorisation, creation/modification/expiry times, etc. The Haggle highlighted two main reasons for this. First, a significant fraction of user data is inherently shared. Secondly, users often carry more than one device. Therefore, even a user's most private data should be network-visible, for transmission to other devices that they own.
6. Build with request-response networking model to avoid relying on a other applications to find the node requested.
7. Exploit all data communication opportunities, including direct connectivity with neighbouring nodes, global connectivity using infrastructure and indirect connectively by exploiting human mobility.
8. Take advantage of brief connection opportunities. Haggle is able to take full advantage of time-limited connection opportunities, by prioritizing potentially exchanged data.

9. Build with resource management. For storage management, Haggle using flags to manage device storage. Haggle proposes networking tasks should be carried out in an order determined by user-level priorities to manage network resources. While Battery resources could be achieved by using context-awareness, by observing the patterns that the user exhibits at various times of the day, the device's location, etc.

Although Haggle was not intended only as an academic exercise in network architecture design, but to be practical and useful, it has not yet been implemented on a large scale. A small prototype of the PSN has been tested in a university project. However, Haggle remain an important step toward achieving a realistic and practical opportunistic networking paradigm for such environment. Further more, this implementation was very useful to understand human mobility pattern in more detail.

In Haggle mobile devices exploit all data transfer methods to communicate including infrastructure networks. While this can be seen as ideal networking capability for mobile devices, it also makes PSN an extension/improvement of current infrastructure networks. The research in this thesis views opportunistic networking as different, by assuming mobile devices can only communicate with each other using infrastructure-less networks. As shown later, this will serve well the potential applications suggested.

2.4.3 Potential Applications

Despite the fact that few applications have been implemented so far, many potential applications will probably follow in the next few years. The OpNet mainly targets applications, where the data to be delivered small, such as SMS or microblogging, and the data itself should be delay tolerant.

During this research, the Arab Spring in Tunisia, Egypt, and Libya started. One of the interesting aspects at that time was that people in Cairo and Benghazi were in desperate need of an infrastructure less model to be able to communicate with each other, simply because the dictators in Egypt and Libya decided to shutdown the infrastructure networks for security reasons. This is a real example where OpNets are a practical solution, along with other emergency situation or natural disaster where the infrastructure has shut down.

Another potential application could be a local advertisement data dissemination application. Often a local shop has a white board where people can put their ads for local potential customers. People usually find it difficult to record the ad details and sometimes do not even see the ads at all. Instead of using a local board, a standalone fixed wireless device could be used so people can store their ads in this device and for those who are interested in this ad, the wireless device pushes this ad to them. Furthermore, this device could enable asynchronous communication between people about any local issues. By their nature, such messages are required to be disseminated in the immediate area.

Other potential application could be an application for spreading word of mouth trust. We often trust word of mouth marketing where satisfied customers tell us how much they liked a business, product, service, or event. While word of mouth remains the best way to learn about the quality things, it has the drawback that if you do not ask, nobody will tell you. Making this problem worse, people are often too shy to ask. Imagine you are in the city looking for a place to eat/buy a service but are really keen to see other people's views, the OpNet platform could easily solve this problem. Assume you want to buy a product today, your mobile device can be set to send a request to all people you pass by asking them if they have reviewed this product before. In the city environment it is easy to get some feedback from others. A week later, you are walking in the street, and another device asks you the same question, so you can pass on to them the information you collected a week ago. In this way, you get plenty of information from people you have never met.

An interested potential application in [3], presented the idea of using an application similar to microblogging services (such as Twitter) among opportunistic networks users. According to their scenario low payload microblogs (utterances) are generated by mobile devices, stored directly in their memory, and opportunistically exchanged upon their pairwise interactions. The forwarding process between mobile devices exploring the role of similarity of interest between users to reduce the number of irrelevant utterances delivered to users. The most interesting aspect of this application is taking the microbloging from the Cyber World to the real life one where the followers are more likely to be your real life friends/followers.

2.5 Data Delivery in OpNets

Data delivery in OpNets is a problem because of two key challenges: the unpredictable mobility pattern of mobile users, and the resource constraints of mobile devices. Therefore, most of the work to date in OpNets is on the design of data delivery protocols.

Data delivery protocols can be classified into data forwarding and data dissemination protocols. This classification is based on the knowledge of the destination address. In a data forwarding protocol, the data that needs to be delivered has a specific destination address, while in a data dissemination protocol, the data objects that needs to be delivered has no specific destination address. Instead of using the physical address, a more meaningful identifier could be used, such as the user name (as used in the Haggle project).

Data forwarding can be defined as the process of moving data from source node to specific destination node. The forwarding process could be one hop or multihop to reach the destination node. Data dissemination can be defined as the process of moving data objects from source node to other nodes who are interested in the data object.

Most of the research on OpNets has focused on forwarding protocols (also known as routing). Because the destination address is known in advance, the main target of the forwarding protocol is to move messages closer and closer to their destinations. Forwarding protocols usually choose the next hop toward the destination by estimating the probability of delivery to the destination node. Information such as node contact history and spatial information can be used to calculate or estimate these values. Although developing forwarding protocols is not the aim of this research, some of these protocols will be highlighted in section 2.5.1 in order to build up our understanding of data delivery techniques and to gain a deeper understanding of opportunistic networking in general.

In data dissemination protocols, the data needs to be delivered to any who are interested, i.e. no specific destination. In this environment, the user generates data and wants to share it with others. The users who generate data are unaware of the nodes interested in their data. While there has been a lot of work on the design of forwarding protocols, little has been done to address the design of data dissemination protocols. A literature review of dissemination protocols will be found in section 2.5.2.

2.5.1 Forwarding Protocols

The literature classifies forwarding protocols, based on the amount of information used in predicting the topological change of the network [64], into flooding-based (also known as dissemination-based) and context-based protocols.

2.5.1.1 Flooding-based protocols

In flooding-based protocols, each node carrying a message forwards that message to all nodes they encounter in order to deliver it to its final destination by a pure flooding policy. The idea behind this policy is that, in the absence of detailed information on a possible path towards the destination, a message should be sent everywhere. If each node forwards the message to every node it meets, it will eventually reach the destination, but with a high overhead of unnecessary message transmissions. Protocols in this class usually yield the highest throughput and the lowest delay but suffer from high cost of energy and resource consumption simply because of the high traffic overhead. The EPidemic Routing (EPR) [82] is an example of such a protocol.

EPidemic Routing (EPR) EPR relies on the theory of epidemic algorithms [26] by passing messages between nodes as they come into contact with each other to eventually deliver messages to their destination. In EPR, each node an index of all messages kept by that node which is called a summary vector. When two nodes meet they exchange summary vectors. The exchange of summary vectors is used to determine if the source node has a message that was previously unseen to the destination. In that case, the node requests the messages from the other node. By applying this method, the messages will spread like an epidemic disease. In EPR, each message must contain a globally unique message ID to determine if it has been previously seen. In addition to source and destination addresses, messages also contain a hop limit field, used to limit the number of hops it is allowed to travel to reach its destination. To limit message flooding in EPR, Network-coding-based routing [85] has been introduced which aims to limit message flooding by combining (encoding) blocks of data together.

2.5.1.2 Context-based protocols

Context-based routing exploits some information that aids in making decisions to identify suitable next hops towards the eventual destinations. Notable examples are the PROPHET protocol [54], which exploits the observed contact history to select the next node. In other protocols, such as HiBOp [9], the node uses more information to choose their next hops. In this protocol, the messages are forwarded to nodes that share increasing amounts of contextual data with the message destination. Other routing protocols, such as MobySpace [43] and MV [15], exploit historical information about nodes mobility patterns and the places nodes frequently visit.

While the above protocols use only mobile nodes to forward messages, others such as

Pan et al. [63] investigate how the possible performance of OpNets can be improved by making use of whatever infrastructure is available. Pan et al. also looked at which conditions will be necessary or useful for successful opportunistic network operation, and how much improvement it can provide over a purely infrastructure-based system. The infrastructure here is the internet access point (AP). Their study concludes that there is a phase transition as an initial deployment of APs had a significant impact on the network performance, but after a certain level, the benefits of additional AP deployments are minor and the use of the opportunistic communication system remains stable. In [75] a similar approach (using infrastructure networks) was presented and shows that opportunistic routing can efficiently utilize infrastructure networks and achieve significantly higher throughput performance.

A different use of fixed infrastructure has been introduced in the Shared Wireless Infostation Model (SWIM) [76]. In the SWIM model there is only one final destination for all nodes. Suppose a node wishes to send a message to the final destination. It will deliver the message to the base station directly if within communication range, otherwise it delivers the message opportunistically to a nearby node that will eventually forward it to the base station when encountered. The base stations, which are distributed across the network region, are gateways to less challenged networks and the goal of the SWIM protocol is to deliver messages to the gateways only. In the approaches given in [63] [75] [76], the use of infrastructure networks is essential, which is not the way an OpNet should work.

2.5.2 Data Dissemination protocols

Research on OpNet protocols has mainly focused on routing/forwarding protocols. Nevertheless, a number of approaches have been developed in recent years focusing on data dissemination. Yoneki et al. [87] present the Socio-Aware protocol for data dissemination, which takes a clustering-based approach. Each cluster/community selects the closeness centrality node as a Broker node. Closeness centrality is a measure of how long it will take data to spread to the others in the community. The closeness of a vertex is the inverse sum of the distances to the other nodes. Socio-aware protocol supports publish/subscribe paradigm which usually contains three elements: a publisher who publishes messages, a subscriber who subscribes his interests to the system, and an event broker network to match and deliver the events to the corresponding subscribers. Therefore when a node has a publication, it sends it to its broker node within the community. When the broker node receives a publication, it matches it against its subscription list. If it matches, it floods the publication within the community. This operation may be performed as either a unicast or a broadcast. The socio-aware protocol needs a lot of traffic to keep up-to-date with who is the broker node, simply because the network topology changes so often. When a broker node changes upon calculation of closeness centrality, the subscription list is transferred from the old one to the new one. Furthermore, the socio-aware protocol fails to deal with situations where the mobile nodes refuse to work as brokers, which is expected to be quite common in an OpNet environment.

Another publish/subscribe paradigm approach is SocialCast presented by Costa et al. [23]. SocialCast is based on the assumption that users with common interests tend to meet each other more often than with other users. SocialCast consists of three phases: Interest Dissemination, Carrier Selection, and Message Dissemination. During Interest Dissemination, each node broadcasts a control message containing the list of its interests to its 1-hop neighbours, along with a corresponding list of utility values. The utility is calculated using social information. During Carrier Selection, this utility is compared, for each interest, against the highest among those communicated by neighbours and if the utility of any neighbour is higher than the local node, it is selected as the carrier. In Message Dissemination, messages are forwarded to the interested nodes and the best carrier. This approach uses interest utility to select the carrier while it is ignorant of nodes willing to be carriers. Nevertheless, this approach is more advanced than the one presented in [87] because it exploits social information in the dissemination.

Boldrini et al. [10] [8] present ContentPlace which is a more refined approach using social information in data dissemination. Specifically, dissemination is driven by the social structure of the network users, so that nodes store data items that are likely to be of interest to users they have social relationships with (and who, therefore, are expected to be in touch with in the near future). Once again, this approach builds on the assumption that a mobile node is usually willing to relay message, which is arguable. Other data dissemination schemes for OpNets include that defined in the PodNet project [51], which does not exploit social information instead the dissemination process distinguishes two phases: during the first phase, the node satisfies the user's requests for the content of subscribed channels. When all available entries of the subscribed channels are downloaded, the device may use the remaining connection time to update and download content that it caches for public use. A different policy is used for caching the public use, such as caching the most popular object, or less popular object or a policy of no cache at all.

Conti et al. [21] present a data dissemination approach using recognition heuristics. In this approach, when two nodes come in contact, they exchange summaries of their data items. Nodes only pull the data that is useful to them, but this approach assumes that each node collaboratively contributes to the diffusion of information by storing and exchanging some of the data they discover coming in contact with other devices, even if those items do not belong to the channel they are interested in. Recognition heuristics are used in selecting data for diffusion processes.

While all the data dissemination protocols mentioned above use only mobile nodes, LeBrun et al. [48] presented Bluespots which is a data dissemination system exploiting standalone infrastructure. Bluespots is a public transport based content distribution system. It does not support any form of peer-to-peer communication, but it is mainly a data distributor for bus riders. Data that is known to be popular is hosted on BCD stations (Bluetooth hubs) which are placed on buses and is made available to bus riders. The lack of peer-to-peer communication in this system restricts the flexibility of what data can be shared and prevents any sort of user generated data to be shared among users.

Wang et al. [84] presented a social-tie-based information dissemination protocol. In this protocol, every mobile node maintains a tie strength table, which records the social tie relationship with its encountered nodes. This protocol including two phases: weak tie-driven forwarding and strong tie-driven forwarding. The idea of using weak ties come from the American sociologist Mark Granovetter who first highlighted the strength of weak ties [32]. In the weak tie-driven forwarding phase, the message first is spread to more weak tie nodes, which leads to an inter-community information spread. Hence, forwarding more tokens to the nodes with more weak ties can increase the spread speed. After a while, the information has been propagated to multiple communities. Hence, the strong ties will play a more important role, which means influential individuals (with many strong ties) can disseminate the information to many individuals in a short period. Therefore, the next stage is a strong-tie driven forwarding, which is like an intra-community information dissemination. One of interesting thing about approach which is despite this popular argument, which is saying that individuals with many strong ties catalyse the information dissemination in society, this paper suggest that individuals with many weak ties are important. However this approach is built around the assumption that all individuals are willing to spread the message to other individual.

2.6 Limitations in proposed approaches

The success of most of the proposed approaches depends on how much the MNs are willing to cooperate in facilitating data dissemination to other mobile nodes, or in other words, the mobile nodes should be fully willing to work as a relay to other MNs. In real life scenarios, MNs usually are not willing to work as intermediate hops for other MNs due to the fact that the MN is carried by humans and they usually set the mobile devices to decline the offer of relay services to other MNs for the following reasons:-

- Security concerns of MNs users, humans usually decline to offer relay services because they often associate this with some security risk. The study [55] shows the type of active malicious behaviour such as supernova and Hypernova could present to OpNets.
- OpNets use resource-constrained mobile devices, users usually try to make an efficient use of device resources. In study [55], some behaviour could also exist with MNs such as free riders where the MN refuses to serve as a relay for other MNs or black holes where the MN drops all of his/her relayed data without forwarding it to other MNs. The MNs with the free rider and black hole behaviour will negatively affect data transmission performance in OpNets. For example, MNs with free riders will require less memory and energy (which is in the user's interest) than MNs without free riders, but the system has to bear the cost in terms of the overall data transmission performance.

For the above reasons, MNs are not the right candidate as the exclusive relays in a practical scenario. In designing a new data dissemination protocol for OpNets, the following considerations must be made.

- Because the MN is willing to relay is not always a guarantee, there needs to be a new approach which is not dependent on MN relay, yet can make use of any available opportunity.
- In OpNets, the users generates data and wants to share it with others (Data dissemination) usually unaware of the nodes interested in their data. Therefore, the proposed protocol should deal with this issue.
- Because of the limitation of communication time in OpNets, pushing the most interesting data to the recipient node is another important role the new approach should take.

- Each piece of generated data has only a limited time to live. After a certain time this data will be invalid, therefore the new approach should only deliver valid data.

Chapter 3

Mobility Model

Evaluation techniques are a key component of successful protocol design. The validation of a newly-developed protocol is a major step towards completing successful protocol design because it enables the researcher to ensure the integrity of, and to find the best settings for, their proposed protocol. This chapter presents the evaluation model that will be used to validate the data dissemination system proposed in later chapters. The evaluation model aims to evaluate performance across a range of metrics. This will be achieved by running an extensive number of tests with different settings and scenarios.

Within the OpNet research community, evaluation of a new network protocol depends almost entirely on simulation. The main point of using simulation is that it allows a network system to be modelled at appropriate levels of detail, while different features of the model can be changed with the corresponding results being analysed. Using simulation enables a researcher to test a wide variety of scenarios at low cost compared to using a real system. It is impractical to conduct large scale evaluation with real devices, and this often results in restricted scenarios (as seen in the Haggle project)

This chapter first discusses some popular simulators used for evaluation of wireless networks. This motivates the design and implementation of “Mobility driven Network Simulator” (MdNS), which is then presented and its main features described. An important component of the MdNS simulator is the underlying mobility model, which aims to mimic the interactions that arise in real human movement. Clearly, the value of the validation of a data dissemination protocol in opportunistic networks is highly dependent on the mobility models used in the simulations, especially if the protocol is designed to exploit information on the movement and relationships of nodes. This is next addressed, where different available mobility models are assessed. Based on this analysis the “Human Mobility Model” (HMM) is proposed as an alternative mobility model that can be applied to evaluate protocols for opportunistic networks. Finally, the proposed mobility model is evaluated, the findings summarized, and a conclusion drawn regarding this contribution.

3.1 OpNets Evaluation Methods

As seen in Chapter 2, opportunistic networks have been the focus of an increasing body of research in the last few years. The main techniques used to evaluate protocols are based on simulation studies, which can be subdivided into those using *synthetic mobility models* as seen in [9], and [88], and those using *real mobility traces* as seen in [41], and [74]. The main problem associated with using a real trace is that those available have usually been collected from very specific scenarios, and therefore their validity is difficult to generalize. On the other hand, the research community is widely agreed that the use of synthetic mobility models is acceptable for general scenarios provided they approximate real traces.

At the start of this research, three of the more popular open source network simulators were investigated for their suitability as testing environments to support this evaluation. These were the network simulator (ns-2) [13], GloMoSim [60], and Scalable Wireless Ad hoc network simulator (JiST/SWANS) [7]. Because of the availability of source code in the open source simulators, making changes in such simulators could be far easier than in commercial simulators such as OPNET simulator [18].

ns-2 simulator is classified as a discrete event simulator [29][28], in which system operations are represented as a sequence of events, where each event occurs at a specific instant in time. ns-2 was originally developed at the University of California Berkeley

[13], and provides substantial support for simulation of TCP, routing, and multicast protocols. GloMoSim is a scalable simulation environment for wireless and wired network systems developed at UCLA (University of California at Los Angeles) [60], employing parallel discrete-event simulation capability. GloMoSim currently supports protocols for a purely wireless network. JiST/SWANS [7] is network simulator composed of two parts. JiST is a high-performance discrete event simulation engine that runs over a standard Java virtual machine developed at Cornell University, and SWANS is a scalable wireless network simulator built atop the JiST platform. The SWANS is organized as independent software components that can be composed to form complete wireless networks or sensor network configurations. JiST/SWANS capabilities are similar to ns2 and GloMoSim, but are able to simulate much larger networks.

From our investigations into these simulators, we noticed that none of these simulators easily supports additional custom mobility models; instead, they rely on random mobility patterns such as random waypoints [44] and random walk [80]. Since node mobility plays an important role for data delivery in OpNets utilising human carriers, the mobility model is an important component of the simulation study. The main role of a mobility model is to mimic the movement of actual users; the accuracy of mobility determines the validity of conclusions. We do not believe that mobility models such as random waypoint and random walk mimic the movement of mobile users with sufficient realism to enable meaningful comparisons.

The main goal of our research is to study the impact of the use of standalone wireless devices to facilitate data dissemination in opportunistic networks, hence we assume perfect wireless links of any wireless technology with no concern about congestion and low-level network performance. The simulators presented in [13], [60], and [7] focus on low-level network performance such as packet throughput and drop rate, while this research requires a higher level performance benchmark, such as average message delivery rate and redundancy. While all three simulators support third-party extensions, investigation revealed that extending an existing simulator would require a considerable number of changes and extensions to these simulators. Therefore, a custom simulator called Mobility driven Network Simulation (MdNS) was developed.

3.2 Mobility driven Network Simulation

From the discussions in Section 3.1, it is clear that a simulator targeting opportunistic networks is really needed. When this simulation was built in early 2008, no network simulator existed that fit the requirements of an opportunistic networks environment. Therefore, MdNS, a Java-based discrete-event simulator was developed. We briefly describe its capabilities and design here.

3.2.1 MdNS components

Figure 3.1 shows MdNS as composed of three main components; the core engine, mobility model, and communication protocol.

3.2.1.1 MdNS Core Engine

In MdNS, simulation operations are defined as sequence of events and the core engine controls the execution and management of these, as seen in Figure 3.1. The core engine initializes system parameters during the "Initialization Subroutine". The system parameters control the simulation clock, number of nodes, type of communication protocol to be used, mobility model setting, and other parameters, which will be discussed in detail in chapter 5. The Event List contains the main operations carried out by MdNS during the simulation run. The possible event types are:

1. Node movement event: The node movement event is responsible for node movement during the lifetime of the simulation test and is controlled by the mobility model used in the MdNS.

2. Node communication event: This event is responsible for communication with other nodes during the simulation time and is controlled by the communication protocol.

3. System update event: This event is responsible for continuously updating the whole system with changes as they happen to allow live visualisation and is controlled by the core engine.

4. Writing log event: This event is responsible for writing logs for all node activities including movement and communication activities during a simulation run. This event is controlled by the engine core.

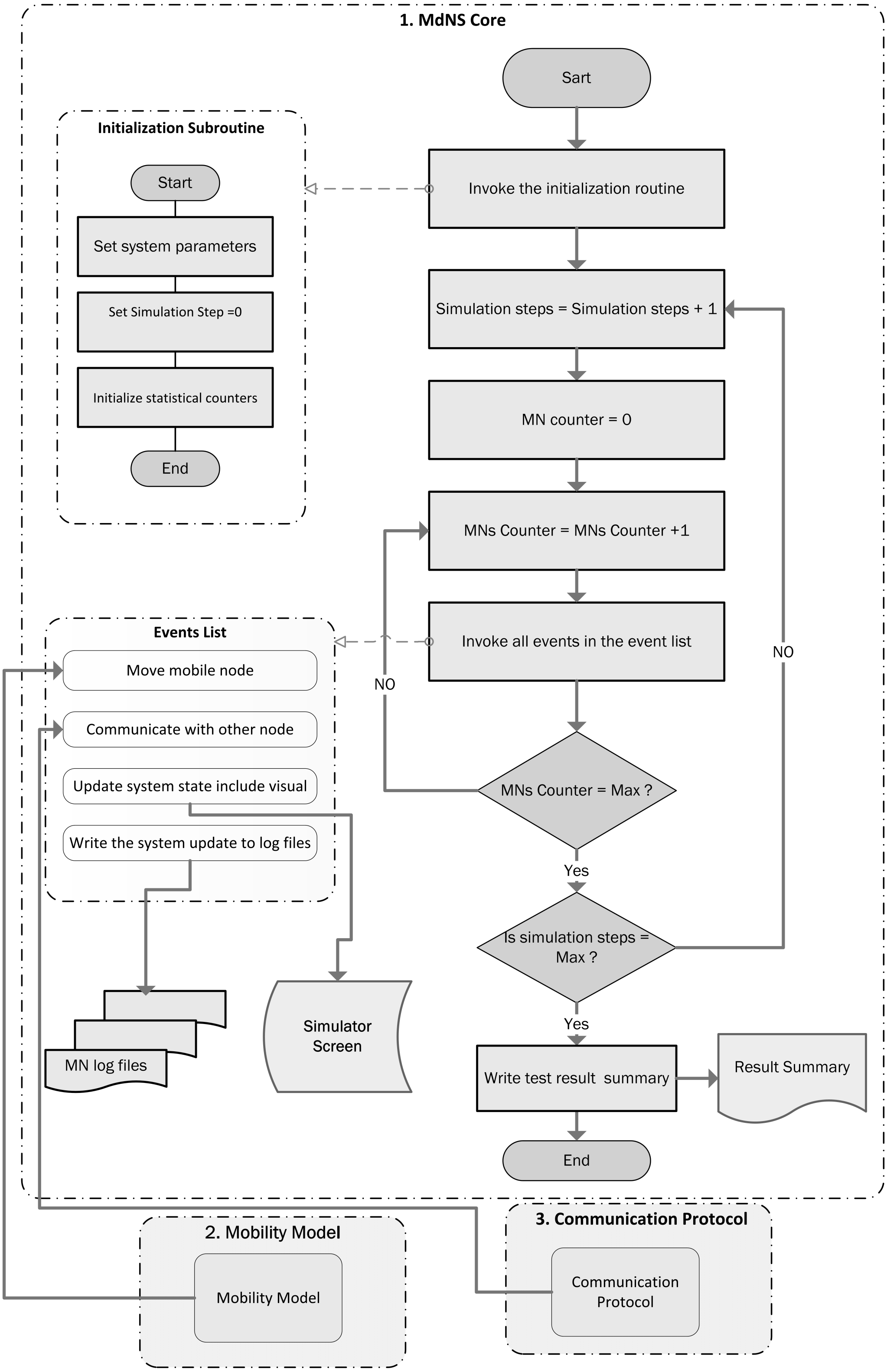


Figure 3.1: Flowchart of MdNS simulation

3.2.1.2 Mobility model

MdNS relies on the mobility model used in the simulation to create the movement patterns for mobile nodes. Details about the mobility model will be presented in section 3.3.

3.2.1.3 Communication Protocol

Within MdNS nodes communicate with each other using the specific communication protocol used in the simulation test. The communication protocol is mainly responsible for which information to push and which information to accept. MdNS implements more than one type of communication protocol. Details about these protocols will present in chapters 4 and 5.

3.2.2 MdNS Features

The main features of MdNS can be summarized as follows:-

1. MdNS uses a text-based script to create the wireless scenarios. In this script, all scenario parameters are defined such as number of nodes, environment size (i.e. the extent of the physical area in which mobile users travel), and communication protocol used.
2. In MdNS, communication payloads are defined as messages rather than packets (as they are in other simulators discussed above) to focus on high level data flow.
3. In MdNS, nodes discover and connect to each other immediately when they are in range of each other; therefore the physical and link layer details are ignored for node discovery. This allows us to focus on the high-level effects of mobility rather than the low-level and make the validation in MdNS not specific to any technology that may implement in a real system.

4. Mobility node speed and wireless range are configurable in MdNS. The wireless range represents the effective wireless range for a node while the mobility node speed represents the speed of mobile nodes, which is used as node speed during the simulation test. Note that the mobile node speed has an inverse relationship with contact duration time for nodes moving in opposite directions.

5. MdNS is capable of displaying the simulation experiment in real time. This display includes node movement, paths, obstacles, and locations. A screen shot of a simulation in action is shown in Figure 3.2. In this figure, the vertices represents the locations and edges represents paths between them.

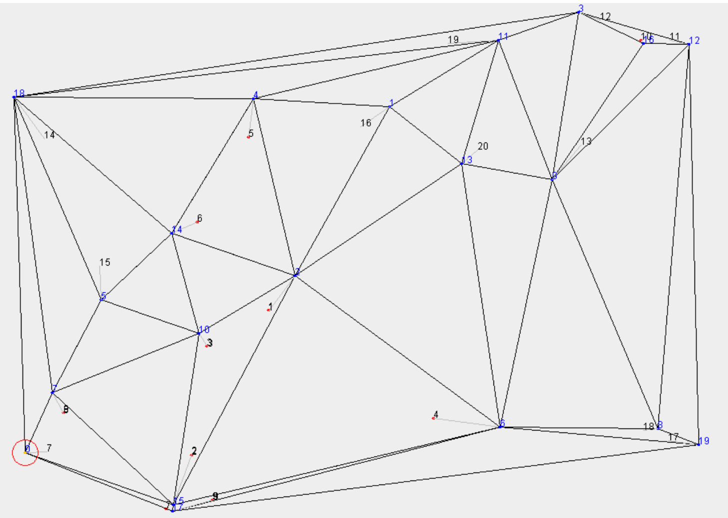


Figure 3.2: MdNS Visual Tool screen shot showing the vertices and paths which constrain movement within the environment.

6. MdNS produces detailed log files at the end of the simulation experiments. These logs include node mobility, contact time, inter-contact time, number of messages ordered as created, sent, and delivered. All of this data is recorded with the time and location at which the event occurred.

3.3 Mobility Models

Node mobility in opportunistic networks contributes strongly to the performance of the data delivery process; therefore a realistic mobility model is a crucial component for the evaluation. Using unrealistic mobility models in the evaluation of opportunistic networks protocols may lead to incomplete or inapplicable results. Due to the nature of the networks studied in this research, where individuals carry their wireless devices, understanding the characteristics of human mobility patterns is needed in order to design a realistic mobility model. The following subsections discuss the characteristics of human mobility patterns, related mobility models, and present the conclusion of the requirements.

3.3.1 Characteristics of human mobility patterns

Understanding human mobility patterns is a key factor in designing realistic mobility models. Some characteristics are obvious, such as the fact that humans usually follow specific paths between origin and destination which are routed to avoid obstacles. To understand human mobility in more detail, a study of real traces of human mobility is needed. Current mobile phone communication technologies provide carriers with detailed information on human mobility across a large percentage of the population. In [73] and [31] this information has been used to study human mobility, while other technologies such as GPS have also been used to study patterns as in [69] and [68].

Gonzalez et al. [31] study trajectories of mobile phone users, based on two data sets, with the first consisting of the movement traces of 100,000 anonymous mobile phone users whose position was tracked for a six-month period. In this study, they used phone activity as tracking tools. Each time a user initiated or received a call or an SMS, the location of the tower was recorded. The second data set contains location traces for 206 mobile users who monitored their location every two hours for whole week. This study concluded that the individuals had a high degree of regularity in their daily travel patterns. They have found that individuals have strong tendency to return to locations visited before within 24 to 72 hours and also they have high return probabilities for a few frequently visited locations.

Williams et al. [86] study the visiting patterns of individuals in three diverse datasets which are a metropolitan transport system, a university campus, and an online location-sharing service. This study concluded that the individuals visit at least one location with near-perfect regularity.

The Levy walk is other human mobility pattern which was first observed in [31] and subsequently in [69] and [68], where a study on human mobility involving 44 volunteers who owned mobile devices equipped with global positioning system (GPS) was conducted. A Levy walk/flight is a statistical description of motion which was introduced by the French mathematician Paul Levy (1886-1971) in 1937 [56]. A Levy flight/walk is a random walk whose step length occurs with a power law frequency.

Reader et al. [67] studied individual route preferences, involving 72 participants from Utrecht University (Netherlands). The study concluded that past experiences and actions of other individuals influence route choice, but if the individuals were aware of a short route, this was usually preferred. The selection of the shortest path route as preferred choice was also revealed in [30] in a study involving 32 participants.

Finally, studies [62],[17] focus on the distribution of inter-contact (ICT) and contact time (CT) within human mobility. The CT is defined as the time when two individuals are located in the same place (therefore their mobile devices are in range of one another and could transfer data if they wished), while the ICT is the time between CT. These studies show that real world human mobility exhibits a strong power law tendency for both CT and ICT. A description of the power law is included in appendix A.

From the studies presented above, the main characteristics of human mobility pattern can be summarised in the following list:-

1. Humans avoid obstacles and move in constrained paths. For each region such as

a city, town, or university campus, people share the same paths or road network.

2. For each human, there is one or more locations where they spend relatively long times without any movement activity. Individual homes and work places are examples of such locations. In this thesis, these locations are referred to as *"off-locations"*.
3. Human mobility patterns have a similar characteristic to Levy flights.
4. Humans choose their destination according to their relationships with different locations.
5. Humans choose their path rationally, usually choosing the shortest path to their selected destination.
6. Temporal and spatial regularity: Humans have a strong tendency to regularly and frequently return to the same location at certain times.
7. ICT and CT between pairs of individuals in the same region follow a power law.

3.3.2 Relevant Mobility models

In the literature, mobility models are classified into two broad categories, *trace-driven* models and *synthetic models* [80], [70]. Trace-driven models use experimentally gathered traces from real users, while synthetic mobility models determine the movement of MNs algorithmically. In this thesis, we are not interested in trace-driven mobility, hence only synthetic models will be discussed here. Synthetic mobility models can be divided into four types according to the driving force of node movements (Figure 3.3), which are: Random, Social driven, Spatial driven and Social-spatial driven. Each of these types is now discussed.

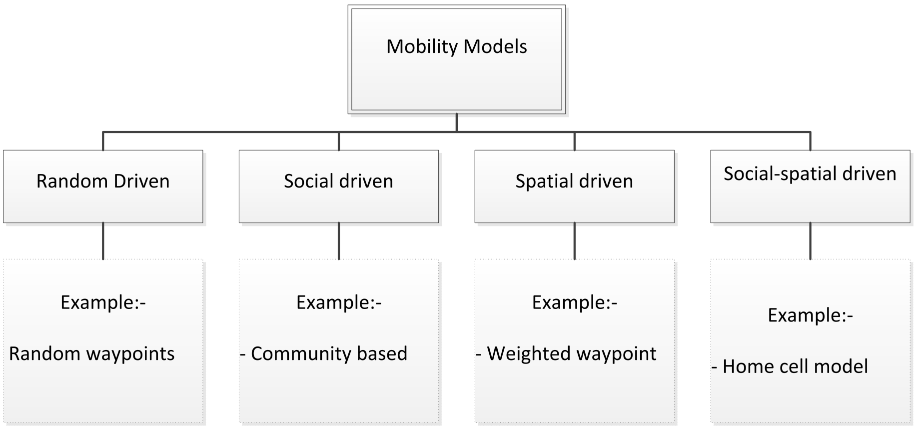


Figure 3.3: Mobility model categories

3.3.2.1 Random Mobility Models

Mobility models under this category share the concept that nodes move randomly without being affected by any social or environmental factors. One of the most widely used and simplest models in this category is the random walk [80]. In this model a MN is initially placed at a random location in the simulation area, and then the MN moves from its current location to a new location by randomly choosing a speed and direction in which to move. The new speed and direction are both chosen from pre-defined ranges, respectively [min-speed, max-speed] and [0, 2π] respectively. At the end of which a new direction and speed are calculated The MN changes its direction and its speed at a constant time interval or when a constant distance has been travelled.. Many derivatives of the Random Walk Mobility Model have been developed, such as Random Walk with wrapping [59]. The objective of this is to address the problem experienced when mobile nodes reach the boundary of their simulation area. In the random walk with wrapping approach, when a MN reaches an edge, it wraps to the opposite edge and continues its movement with the same direction and speed. Although the Random Walk is the simplest mobility model the researcher can implement, it generates unrealistic movements with respect to human mobility, such as sudden and sharp turns.

Another popular random driven mobility model is random waypoint (RWP) [44]. In

RWP model, each node is initially placed at a random position within in the simulation area. As the simulation progresses, each node pauses at its current location for a period of time possibly zero, and then randomly chooses a new location to move to in straight lines. Each node continues this behaviour, alternately pausing and moving to a new location for the duration of the simulation. One of main difference between Random walk and RWP is that RWP introduced pauses between changes in direction or speed. A deeper study [50] investigates the effects of RWP and random walk mobility models on the opportunities for peer to peer data exchange, concluding that the RWP model also provides substantially fewer opportunities for exchange compared to random walk. However, this model cannot be considered as suitable for simulating human movement,due to the continuous changes of direction that characterize it.

All the models mentioned so far are based on the same assumption that MNs have the ability to move freely without regard to any environmental obstacles surrounding those nodes (e.g. building, walls, etc.). For most real life scenarios, it is unrealistic to assume that the nodes are allowed to move across the entire simulation area due to the existence of obstacles. Therefore these mobility models fail to comply with the first and second characteristics of human mobility patterns presented in section 3.3.1. In order to comply with these characteristics, Tain et al. in [79] present a graph-based mobility model. In this model a graph is constructed with vertices modelling locations that the users might visit and edges modelling the connections between these locations. Figure 3.4 shows vertices which represent locations and edges representing the paths possible between these locations.

In the graph mobility model, a mobile node is initialized at a random chosen vertex in the graph and then their destination is chosen randomly. Mobile nodes use the shortest path to reach their destination. After reaching their destination, nodes pause for a random time and then pick another destination randomly and moves to this destination. This process continues until the simulation time ends.

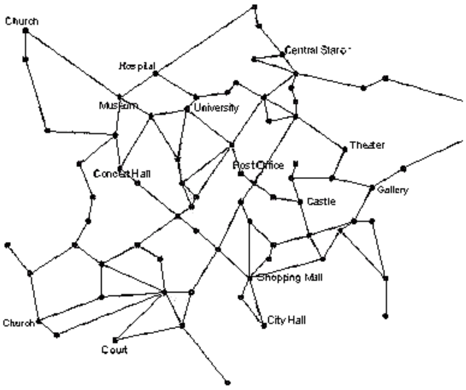


Figure 3.4: An example graph modelling a city center from [79]

It's clear that this approach complies with the first characteristic of human mobility pattern "Avoid obstacles and move in constrained paths". Using the shortest path to reach their destination makes this approach complies also with the fifth characteristic "Choosing the shortest path to reach their selected destination". However this approach fails to comply with the remaining characteristics of human mobility patterns. Because no off-location is associated with mobile nodes, the model fails to comply with the second characteristic. The uniform randomness with which nodes choose destinations makes this approach fail to comply with the third, fourth, and sixth characteristics of human mobility pattern.

3.3.2.2 Social Driven Mobility Models

Social ties between nodes are the main driver of node movement for these mobility models. An example of such a model is the Community based Mobility Model (CMM) as seen in [58], [36]. In CMM nodes are grouped into social communities. Nodes that are in the same social community are called friends, while nodes in different communities are called non-friends. Relationships between nodes are modelled through social links. Social links are used in CMM to drive node movements. Nodes move in a grid, and each community is initially randomly placed in a cell of the grid. Each node is initially associated to a certain cell in the grid. When the model is initially established, a destination for each node is randomly chosen inside the cell associated to its community . When a destination is reached, a new destination is chosen according to the social attraction exerted by each cell on the node. Attraction is measured as the sum of the link weights between the node and the nodes currently moving in or towards the cell. CMM fails to comply with the first characteristic “ human avoid obstacles and move in constrained path. Furthermore, Boldrini et al in [11] highlighted other problems associated with CMM, showing a gregarious behaviour, where all users in a community tend to follow the first user that moves outside the physical location where the community is located.

The mobility models presented in [12], [36] also fall in this class. In [36] the Social Mobility Model is proposed, in which a node periodically moves between abstract locations (called anchors). An input network specifies the relationships among the users, creates these anchors and allocates a set of anchors to each node. In this model, a node has to visit all of their anchors in every round of the simulation time. At the beginning of the next round, the node starts to cycle through its anchors again, as a simple model of scheduled and periodic meetings with acquaintances. A more advanced model can be seen in [12], which proposes Sociological Interaction Mobility for Population Simulation (SIMPS). In the SIMPS model, nodes move according to two behavioural rules, namely *socialize* and *isolate*. The *socialize* behaviour rule enables a node to move towards acquaintances, while the *isolate* rule is designed to enable a node to escape from undesired presences, i.e. strangers. These two behaviour rules balance the volume of current social interactions against the volume of interactions needed by the node. The SIMPS model divided into two parts: social motion influence and motion execution unit. The social motion influence updates an individual’s current behaviour to either socialize or isolate. The motion execution unit is responsible for translating the behaviour adopted by an individual into motion.

3.3.2.3 Spatial Driven Mobility Models

The mobility models mentioned above do not address spatial preferences within an individual's chosen destinations. This is an important issue because humans do not choose destinations purely randomly. Hsu et al. in [37] proposes the Weighted Way Point (WWP) mobility model. The major differences of WWP model and the popular RWP model are that mobile nodes no longer uniformly randomly choose their destinations. Instead the weights for choosing the next destination depends on both the current location and time, and the pause time distribution at each location is a property of that location.

One of the most flexible models among those in this class is the Time-Variant Community model (TVC) [38], a model proposed to capture the realistic mobility characteristics the authors observed from wireless LAN user traces. Specifically, communities (areas that a user visits with high probability) are used to model the highly skewed preference for visiting locations and time periods with different mobility. Parameters are used to model the periodical reappearance of a node at the same location. These two properties have been observed from multiple WLAN traces and serve as the motivation for the proposal of such a mobility model. In TVC, each user can be either in a normal movement period or in a concentration period. In each period a node is assigned to a (different) community, that represents the frequently visited location for that node. Then, each node moves within its community or across the whole network according to a binary-state. Different locations (here, communities) realistically can have different popularity, e.g. at different times of the day (normal movements or concentration period). While the two mobility models WWP and TVC deals with time and space preference of the users, they fail to deal with the social aspect of human mobility pattern.

3.3.2.4 Social-Spatial Driven Mobility models

A social driven mobility model such as CMM is usually not able to capture the attraction applied to users by physical locations such as their home locations. In [11] Boldrini et al extend the CMM model to include attraction exerted by physical places. The authors introduced Home Cell Mobility Model (HCMM) which can be considered as an extended version for CMM (discussed above). HCMM, joins the concepts of CMM (for modelling social relationships between users) with the concept of defining preferential locations in which users tend to spend most of their time. In HCMM, when a node is in its home cell, the cell for the next movement is selected as in CMM. However, after a node reaches a cell which is not its home it then return to its home. While HCMM was able to comply with some important characteristics of human mobility, such as the second characteristic where each human has a attraction for specific location such as it's home, this model fails to comply with other important characteristics such as the first characteristic " human avoid obstacles and move in constrained path".

3.3.3 Summary of mobility models

Most mobility models mentioned above do not comply with all important characteristics of human mobility patterns relevant to this study. While some of these mobility models satisfy a number of these characteristics, they ignore other important ones. For the purpose of evaluation in this research, a realistic mobility model which complies with the important characteristics of human mobility pattern is needed. The following list summarized the important characteristics of human mobility pattern which needs to be included in the mobility model in order to mimic human mobility in realistic way.

- Humans move in constrained paths: Due to this character, humans are able to share the same paths as result they are able to communicate more often. Therefore the mobility model should comply with character.

- Humans choose their destination according to their preferences: Because of this character, human mobility characterize by spatial and time regularity as seen in [31][86]. In our research we employed this regularity therefore mimicking this character is an essential for the purpose of our evaluation.
- Humans often choose the shortest path as seen in [67][30]. Because of this character some location/path are often used more than other. In our research we have employed this character so this character should be included in the purposed mobility model.
- Humans have a strong tendency to frequently return to the same location as seen [86]. This character also affect the temporal and spatial regularity of human mobility which we have employed in our research therefore it an essential character in the proposed mobility model.

3.4 Human mobility model

As the existing mobility models fail to comply with all of the relevant characteristics of human mobility pattern to this research, the need for a new mobility model becomes an essential step. In this section, we introduce the *Human Mobility Model* (HMM). Building process of the HMM is divided into three main phases, the design, the implementation, and the validation. In the following subsection we present these phases in detail.

3.4.1 HMM Design

The main objective of this mobility model is to mimic real movement of humans yet to maintain a small number of system parameters to allow evaluation of our system under different conditions. In our design, we have divided the design process into two parts. First, the environment model representing elements such as paths, locations and obstacles. Secondly, the movement model which identifies the rules for updating the position of mobile nodes.

3.4.1.1 Environment Model

In order to mimic the real life environment where a human has paths to follow and obstacles to avoid, we use a synthetic map so we can easily experiment different environments. To model the environment a graph $G(V, E)$ is defined, with V representing discrete locations in the regions and edges E representing direct routes between these (Figure 3.5). For each edge $e \in E$ a weight $w(e)$ is defined representing the distance between vertices. By using distance to represent weights, G can be defined as a uni-directional weighted graph. The number of edges connected to vertex v is defined as the degree of a vertex, $g(v)$. Attention is restricted to connected graphs, i.e. at least one path exists between each pair of vertices hence $g(v) > 0 \, \forall v \in V$. In the simulation, we assume V is composed of two disjoint, non-empty subsets of vertices named *Junctions* (J) and *Off-Locations* (F), so that

$$V = J \cup F \, \text{and} \, J \cap F = \emptyset \tag{3.1}$$

With loss of generality in this thesis, for simplicity, we treat J and F as disjoint subsets. In practical, locations may be in both sets. For example, park locations maybe considered as destination and junction as compared to lecture theatre which is only a destination.

A *junction* models a location where more than one route is available, and the MN can make a decision on which route should be taken, hence vertices in J satisfy:

$$g(v) \geq 2 \quad \forall v \in J \tag{3.2}$$

An *off-location* models a venue where a MN may spend some time, such as a home, work place, coffee shop, etc. This location cannot be used as a hop to reach other destinations, hence vertices in F satisfy:

$$g(v) = 1 \quad \forall v \in F \tag{3.3}$$

M is defined as set of mobile nodes $M = \{m_1, m_2, m_3, ..., m_{N_m}\}$ which travel through the environment defined by G, each representing a mobile handhold device equipped with short wireless technology carried by humans. Each $m \in M$ has a vertex $v \in V$ which is considered to be it's home location and denoted as $h(m)$. In Figure 3.4 $J = \{v_1, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{12}, v_{13}, v_{15}, v_{16}, v_{17}\}$ and $F = \{v_2, v_{10}, v_{11}, v_{14}\}$.

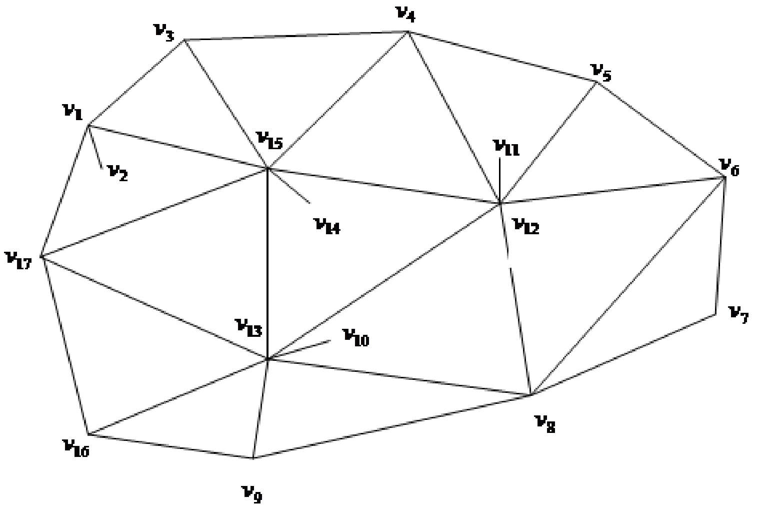


Figure 3.5: Example of Graph G(V,E) used as map

3.4.1.2 Movement Model

The movement of a mobile node m through the environment is specified as follows:-

1. Define a spatial preference relationship for m: Suppose there are at least two off-locations ($|F| \geq 2$). For a given off-location $f \in F$, we define the spatial relationship, $SpR(m, f)$ between a mobile node m and the destination f as the probability that m will choose f as its next destination if m is currently located

at $h(m)$. Thus

$$SpR(m, h(m)) = 0 \quad \forall m \in M \tag{3.4}$$

$$\sum_{f \in F} SpR(m, f) = 1 \quad \forall m \in M \tag{3.5}$$

2. Choosing a path for m: Since G is a connected graph, at least one path exists between any two vertices in G. A path P is referred to by the sequence of its vertices; say $P =< v_0, v_1, ..., v_k >$ means P is a path from v_0 to v_k. In this mobility model, m chooses a shortest path to reach destination and if multiple shortest paths exist, m selects one at random.

3. Movement regulation: At the start of simulation, assume m is placed at its $h(m)$. After specified time which refereed to as "*off-time*", m will choose its next destination according to SpR. m reaches its destination by following the shortest path in the graph with a constant speed. If the node arrives at its destination, it pauses for some specified time and then returns to its home location by the reverse path.

3.4.2 HMM Implementation

This section will explain the implementation of the mobility model in detail. Implementation of HMM, mirroring its design steps, is organized as follows: First, the environment model is created. This section explains which graph design was used to meet the requirement. Secondly, the movement model is defined, where the spatial and social relationship is defined and which shortest path algorithm has been implemented.

3.4.2.1 Simulating Environment Model

As stated above, HMM uses a graph for representing discrete locations and connected paths between them. In our implementation, a graph is constructed by distributing a chosen number of vertices with specific x,y coordinates and edges added to create their Delaunay triangulation [19]. Delaunay triangulation of a vertex set is a triangulation of the vertex set with the property that no vertex lies inside the circumscribing circle (circle that passes through all three vertices) of any triangle in the triangulation. Further information about how Delaunay triangulations can be created are described in Appendix B.

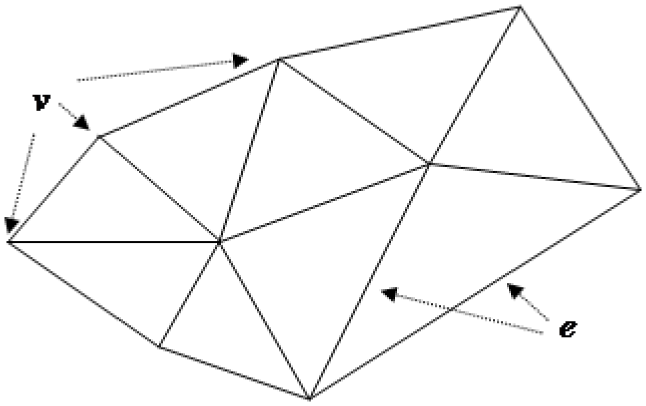


Figure 3.6: 9 vertices connected by Delaunay triangulation

Use of Delaunay triangulation produces a graph $G(V, E)$ as seen in Figure 3.6, where V are vertices and E edges connect them. The weights for the edges are the Euclidean distance between the vertices. The resulting graph G is composed from a set of triangles.

Modelling *off-locations:* As we stated in HMM design section, we assume G is composed of two disjoint, non-empty subsets of vertices named *Junctions* (J) and *Off-Locations* (F), so that

$$V = J \cup F \,\text{and}\, J \cap F = \emptyset \tag{3.6}$$

In this thesis, we assume off-locations and junctions are disjoint. Note that this requirement could be relaxed without any loss of generality.

For a mobile node m, the path to any selected destination should include a node $v \in J$, therefore if the off-locations are placed in J, then m would encounter other nodes at their off-locations on route toward the selected destination and consequently communicate with them without intending to do so. Hence we strongly believe that having the off-locations in the $v \in J$ negatively effects the accuracy of the result. For this reason, in modelling the off-location F for HMM this point was taken into account.

For a triangle within G with vertices A, B, and C, denoted ΔABC. Each vertex in ΔABC has x and y coordinates, i.e for vertex A, the coordinates are A_x, A_y, vertex B the coordinates are B_x, B_y and vertex C, the coordinates are C_x, C_y. In ΔABC, the edges connecting these vertices are e_{AB}, e_{BC}, and e_{AC}. In ΔABC, we create potential *off-locations* by the following geometric construction:-

1. First, we calculate the coordinates for a middle point k in each edge, i.e. for e_{AB}, k_{AB} is defined as
$$k_{AB} = (\frac{A_x + B_x}{2}, \frac{A_y + B_y}{2}) \tag{3.7}$$
2. Second, we connect straight lines between vertices to their opposite middle points , i.e. $line(A, k_{BC})$ a line connecting vertex A and k_{BC}, $line(B, k_{AC})$ a line connecting vertex B and k_{AC}, and $line(C, k_{AB})$ a line connecting vertex C and k_{AB}.
3. Third, we can allocate three off-locations for ΔABC as follow for the first off-location f_0 located at 20% of the $Line(A, k_{BC})$ and $Line(f_0, A)$ is a line connecting the f_0 to A. The second off-location f_1 located at 20% of the $line(B, k_{AC})$ and $Line(f_1, B)$ is a line connecting the f_1 to B. The third off-location f_2 located at 20% of the $line(C, k_{AB})$ and $Line(f_2, C)$ is a line connecting the f_2 to C. Figure 3.7 shows a graph G with off-locations.

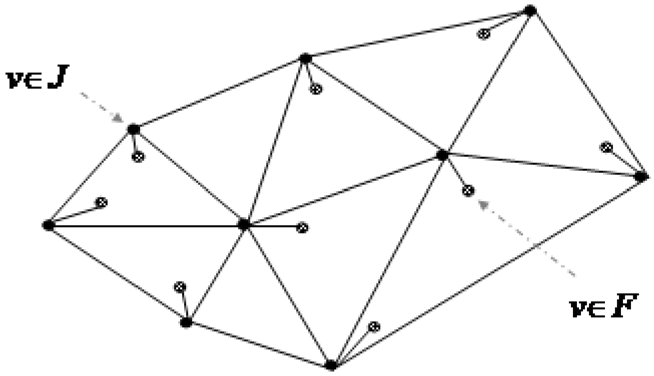


Figure 3.7: Off-Locations Vertices

Let $H(v)$ be the number of off-locations connected to v where $v \in J$. By applying the same rules for created off-location for all triangles, $H(v) \geq 1 \; \forall \; v \in J$. In our simulation $H(v)$ is configurable between 1 to g(v).

Figure 3.7 shows the two types of vertices (F and J) and as can be seen, each off location vertex $v \in F$ is incident to only one vertex $v \in J$, therefore $g(v) = 1 \; \forall v \in F$ and $g(v) > 1 \; \forall v \in J$. As seen in Figure 3.7, G clearly defines the paths between vertices which mobile nodes follow when moving and all the areas outside these paths are defined as obstacles that mobile nodes will avoid.

3.4.2.2 Simulating Movement Model

This section describes how the movement model is implemented, which includes how a mobile node chooses its destination, how it chooses its path and finally how movement happens.

Choosing Destination As stated above users choose their destination according to their personal preferences. Users generally have a special relationship with certain locations, for example their home, work, etc. we assume that the preference each node has for an off location can be expressed as a probability, and further after that each time a node visits a destination, it returns to its home. In the experimental scenarios in this thesis, probabilities *SpR* are defined using Zipf's distribution power law [33] in order to model a distribution ranking (a description of the power law is included in appendix A). Zipf's law states that the size of the largest occurrence of an event z_i is inversely proportional to its rank and is defined by Equation:

$$z_i \propto \frac{1}{i^{\alpha}}(i = 1, ..., N) \tag{3.8}$$

where α is close to unity and N is the number of distinct occurrences of the event.

Although Zipf's law was initially used in the context of word frequencies, it has been found to be a good approximation for many other contexts [57]. Zipf's law states that the most frequent word will occur approximately twice as often as the second most frequent word, which occurs twice as often as the fourth most frequent word, etc. In the same way, humans usually have a strong relation to a few locations such as home, while it has a weaker relation to other locations such as a far-off shop.

Assume a graph G has n vertices V = {v$_0$, v$_1$, v$_2$, v$_3$, ..,v$_{n-1}$}. Applying Zipf's law with $\alpha = 1$ for a mobile node m with home v_0 to these vertices gives $z_1 = \frac{1}{1}$, $z_2 = \frac{1}{2}$, $z_3 = \frac{1}{3}$, etc. Hence

$$SpR(m, v_i) = \frac{z_i}{\sum_{j=1}^{N} z_j} \tag{3.9}$$

Choosing the Path A primary function of the proposed mobility model is to find paths between any pair of vertices within the simulation. As G is a connected graph, at least one path between any two vertices exists. In the mobility model, m choose the shortest path to reach the chosen destination. Dijkstra's algorithm [45] was used to calculate the shortest path between vertices. In the case where more than one path has the same length, the mobility model randomly picks one of them.

Movement Rules The mobility for a mobile node m is guided by the following rules:

1. At the beginning of the simulation, each m will be placed at their home location $h(m)$.
2. Once the simulation starts, m pauses for some time (the duration of time chosen uniformly randomly within predefine time range) before starting to move. This pause time as the start is defined to mimic the reality of human mobility pattern where individuals do not usually start moving at the same time.
3. After a mobile node pause finishes, m will choose a destination $v \in F$ according to their SpR values using roulette wheel selection [91].
4. m starts moving toward their chosen destinations following the shortest path. In each step of the simulation, m moves one distance unit following the graph edges toward their destination.
5. If m arrives at its destination, it pauses for some time (the duration of time chosen uniformly randomly within specific time range) and then returns to its home location by the reverse path following process in step 3.
6. m repeats this process until the end of the simulation time.

3.4.3 Validation of the HMM model

In order to validate this model a number of tests have been carried out. In particular, properties of real mobility traces collected by Intel Research Laboratory in Cambridge [16] have been compared with the trace generated by this mobility model. This section presents and discusses the results of these tests.

3.4.3.1 Validation metrics

Within the research community, it is widely agreed that more contacts between mobile nodes usually leads to more data delivery. As we discussed in Section 3.1, many studies

[62], [17], and [16] show that ICT and the CT of human mobility exhibit a strong power-law tendency therefore, ICT and CT were used as the main metrics to validate our mobility model, similar to the validation strategy used in [58].

3.4.3.2 Simulation setting

Because the synthetically generated traces are being compared with real traces, the test setting was built as close as possible to the real traces. The real traces was collected using 9 mobile nodes and 1 fixed nodes, therefore the synthetic traces was also generated using the same number and types of nodes.

To get a more accurate result, three synthetic traces were created using three different environment maps which place vertices based on different random seeds while the same movement model was kept for all three maps. Table 3.2 shows the simulation settings.

Parameters	Value
Duration	86400 steps
$\lvert J\rvert$	15
$\lvert F\rvert$	10
Minimum off time	7200 steps
Maximum off time	21600 steps

Table 3.1: The test setting used to create three different synthetic traces

A comparison of the properties of synthetic traces (three samples) generated by HMM with those of three real traces (Intel research laboratory, computer laboratory and Infocom) provided by Intel Research Laboratory in Cambridge [16][62] are presented in Figures 3.8 and 3.9.

The comparison between the synthetic traces generated by the HMM and the real traces has been conducted. Figure 3.8 shows contact time (CT) cumulative distribution for both HMM and real traces and Figure 3.9 shows the inter-contact time (ICT) cumulative distributions. In both figures the cumulative distributions presented using log-log coordinates. Figure 3.8 and 3.9 shows these distributions are extracted from the real and synthetic traces generated by HMM shows the approximate power law behaviour for a large range of values like those extracted from real traces. As stated in section 3.1, that ICT and CT for real human mobility pattern exhibits a strong power law tendency and this also can be seen the presented results.

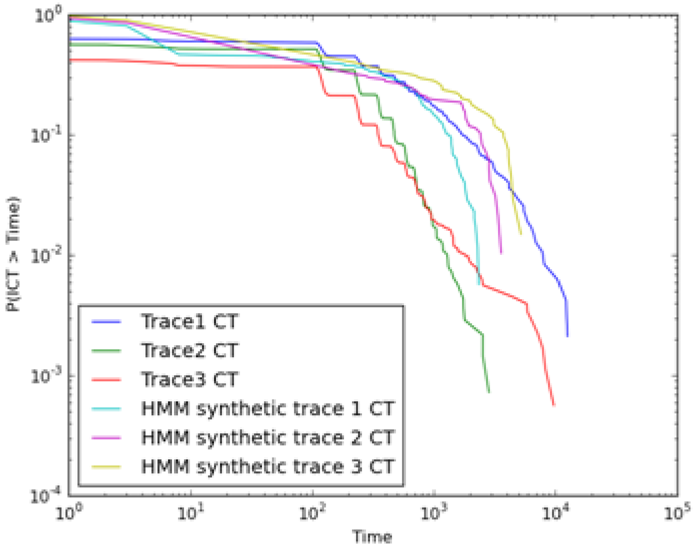


Figure 3.8: CT Test

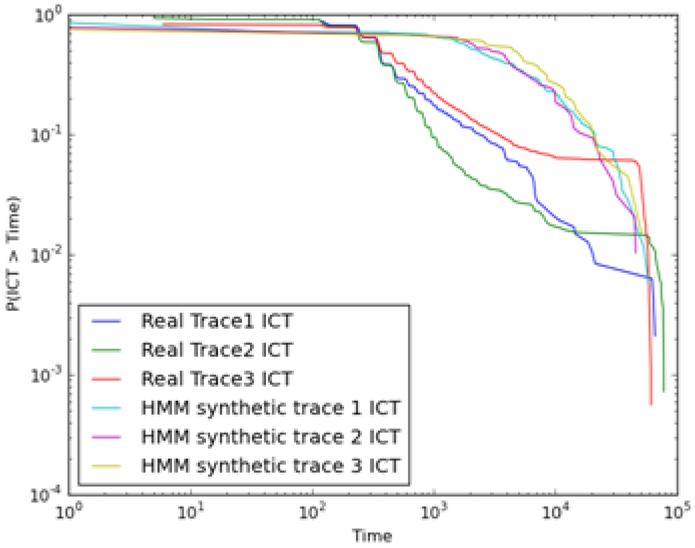


Figure 3.9: ICT Test

3.5 Conclusion

In opportunistic networks, the mobility of users plays a key role in data delivery. Thus, providing an accurate mobility model capable of capturing the key characteristics of human mobility patterns is an important item to evaluate data dissemination strategies in such networks. Many studies presented in this chapter were carried out to understand human mobility patterns and their main characteristics highlighted. From mapping these characteristics to the mobility models discussed in this chapter, we concluded that these models fail to comply with all of these characteristics. Therefore, we proposed a new model combining those features which appear to capture real world characteristics. Evaluation of this model shows that it exhibits well-known statistical features of real human mobility traces.

Chapter 4

Repository based Data Dissemination System

This chapter describes the Repository based Data Dissemination system (RDD) for OpNets. In Section 4.1, we introduce RDD objectives, then we present the design assumptions in Section 4.2. In Section 4.3, we present the RDD design. A summary of the chapter is presented in Section 4.4.

4.1 Assumptions and objectives

RDD is a networking system aiming to provide practical and decentralized push based data dissemination for OpNets. We have set up three main assumptions to achieve our goal, which are as follows:-

4.1.1 Data dissemination without requiring MN cooperation

Most data delivery approaches seen in Chapter 2 are based on the assumption that mobile nodes are fully willing to act as relays. However, that is not the case in real life where mobile nodes are not likely to be fully prepared to work as a relay. Therefore building a system with the assumption that MNs are fully willing to act as relays is not a practical solution. Based on this fact, in RDD relaying data objects to other mobile nodes is not solely reliant on mobile nodes cooperation. Instead, RDD uses carefully placed repositories to relay data objects.

In this thesis, a repository is defined as a standalone fixed wireless device with no access to fixed infrastructure which use a short wireless technology such as ZigBee or Wifi to communicate. Although adding these repositories to the network will change the definition of OpNets introduced in chapter 2, these repositories are just acting like a person standing still and are not acting as network infrastructure in the traditional sense. Although this thesis has not implemented a prototype, the repository envisioned in the model is practical, as a small wireless device, with the following characteristics:

1. It is a completely standalone wireless device and it is distinct from traditional access points which are connected to the central network infrastructure.
2. The repository can easily be installed and moved.
3. Could be similar in hardware, size, and cost to a wireless device prototype "Intel iMote" (Figure 4.1), which was used in [16]. The iMote has been used to collect connection opportunity data and mobility statistics by some researchers at the University of Cambridge. The iMote has an ARM processor, Bluetooth radio, flash RAM, and CR2 battery. The estimated cost of such a device should be less than £100. Because the iMote is a compact device, it will be easy to install and remove.

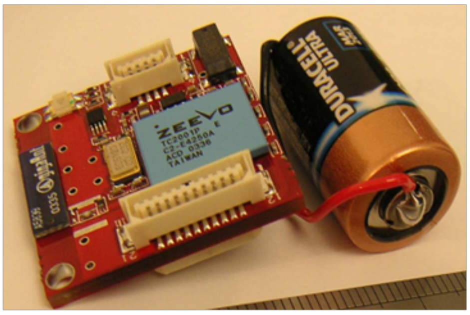


Figure 4.1: Intel iMote device connected to a CR2 battery from [16] showing the potential size of a repository.

4. The repository could use an AC power supply to eliminate the power constraint.

4.1.2 Data dissemination employ mobility characteristics

In OpNets, mobile nodes represent two integrated parts which are the mobile device and its user (as seen in Figure 4.2). As such the data delivery process in OpNets exploits the wireless communication of mobile devices as well as the user mobility; hence the characteristics of the carriers mobility plays an important role in the data delivery. While the majority of existing protocols have focused on the development of the mobile devices themselves, user mobility pattern characteristics have not received the same attention. Since the users' mobility is characterised with a high degree of temporal and spatial regularity where encounters between peers exhibit regular structure. Based on this fact, in RDD we utilize these characteristics to find the most commonly used routes or locations in the target environment to place the repository, to work most effectively as relay nodes.

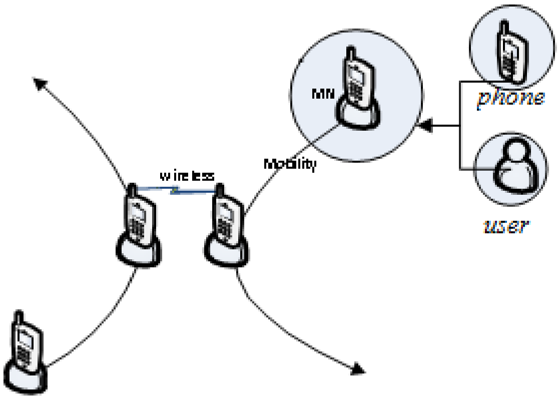


Figure 4.2: Two integrated parts of MNs in OpNets

4.1.3 Decentralised push based data dissemination

Although the mobile nodes in OpNets are independent identity of nature, many protocols use a clustering-based approach where a special node in the cluster plays the major role in the dissemination process between other nodes as seen in [87]. In RDD the nodes are completely independent in which data objects to disseminate and which other devices it disseminate to. In the push based dissemination system, nodes neither advertise their data object's availability nor send request for their needs to other nodes instead RDD employs only nodes preferences to deliver data objects to their interested users.

4.2 Environment assumptions

To describe the simplest form of the protocol, suppose a mobile node s wishes to send a data object to a set of other mobile nodes who are interested in this message. If s comes directly within the communication range of one of the interested nodes, it will deliver the data object directly to that node. Otherwise, if s comes within communication range of a repository r, it delivers the a copy of data object to r , which will eventually forward this message to interested nodes if encountered. The assumptions underlining the design of RDD as follows:-

1. Mobile nodes are carried by individuals walking in environments constrained by obstacles, such as a city. This implies that there are areas more commonly visited
2. Mobile nodes are not aware of their absolute geographical location. This assumption is based on currently available location technologies, where keeping track of the position of a device (for example, via GPS) requires a prohibitively high cost in battery resources.
3. Mobile nodes do not maintain lists of other nodes or their preferences, i.e. apart from during encounters, each mobile node is completely unaware of who requires individual data objects and this is not stored for later use.
4. The target environment of the network is a university campus size or a region of a city. For larger environments, the protocol needs to deal with other issues, shown in chapter 6 when this assumption is relaxed.

5. Data objects are relevant over the period of a single day and we assumed that the buffer size is unlimited.

4.3 Design

RDD is composed of four parts. The first part is the data objects management, which responsible for managing data objects. The second part is user preferences management, which is responsible for managing user's subscriptions to channels. The third part is the data dissemination, which is responsible for pushing data objects to intended mobile devices. The fourth part is the repository placement, which is responsible for placing the repository in the most suitable place. In the following subsections, we will discuss these parts in details.

4.3.1 Data objects management

The mobile nodes form the backbone for the OpNet. At first part, we present the characteristics and nature of mobile nodes used in RDD and then we describe how mobile nodes manage data objects and channel subscription.

4.3.1.1 Mobile nodes

Mobile nodes in OpNet are composed of two parts: the mobile wireless device itself, and the carrier of the device. The nature and characteristics of these parts are now discussed in detail.

Wireless devices The type of mobile wireless device envisaged in this study is a typical smartphone or tablet with short range wireless radio connectivity. In this research, the approach is not restricted by the type of wireless technology used, however all devices are assumed to be equipped with the same technology, i.e. a homogeneous network. This technology could be Bluetooth, ZigBee, WiFi, or any other short range wireless technology operating in unlicensed bands. The mobile wireless devices are usually small. The device size, portable nature, and user lifestyle, place a number of constraints on these devices. Managing these constraints is the major challenge in designing and implementing data dissemination protocols for opportunistic networks. The following points summarize these constraints.

1. Limited memory storage: Most mobile devices have limited storage, which has a strong impact on their capability to store data on behalf of other mobile nodes.
2. Limited battery life: Batteries are also important for the mobility and portability of mobile devices. Batteries run down quickly in most handheld devices. Therefore adding traditional opportunistic networking where devices are required to work as relays for other nodes will add extra load on the batteries of these devices.
3. Limited wireless range: The nature of the wireless technology used in opportunistic networks is a short range, usually stretching from 10 to 100 metres.

Device carriers In the definition of opportunistic networks used in chapter 2, the mobile devices are carried by people, therefore the device carrier has all the mobility characteristics of human movement. Because opportunistic networks exclusively employ device mobility, in addition to wireless media, to achieve data communication between devices, the carrier's movement plays an important role in data communications. All mobility characteristics of human movement which were studied in chapter 3 are applicable here. In RDD, we exploit these characteristics in choosing the repository location. The main mobility characteristics can be summarized as:

1. The carrier mobility patterns show a high degree of temporal and spatial regularity.

2. The carrier chooses destinations based on their personal preferences.
3. The carrier takes the shortest route to reach their destinations.
4. The carrier travels along selected pathways meaning they have the ability to avoid environmental obstacles and they share the paths/locations with other.

4.3.1.2 Managing Data Objects

In our design, data objects are of any kind which are small in size, such as a short text messages, small pictures or video clips. One of interesting aspect of smartphones nowadays is the ability to create such data objects in an easy way. In RDD, we assume that most data objects injected into the networks are created by individual mobile nodes, however the data objects also could be generated by other parties but still the mobile nodes are the only parties responsible for injecting the data objects into the network.

Each node maintains a buffer consisting of all data that it holds. When a mobile node *m* creates a new data object, this data object is added into their buffer. We assume that each data object fit in one packet. Each data object packet d is composed of two parts: the data to be delivered (or ‘payload’); and a header composed of 5 fields as follows:

1. $c(d)$: Channel of the data object d which classifies the content and will be discussed in detail in the next section.
2. $m(d)$: Source node Id.
3. $Id(d)$: Message sequence number, as created by the source node. Hence each message is defined by a unique pair <source node Id, message Id>.
4. $ct(d)$: Data object creation time.
5. $ttl(d)$: Time to life for data object d.

Header					Payload
$Id(d)$	$c(d)$	$m(d)$	$ct(d)$	$ttl(d)$	Data object content

Table 4.1: Packet structure for a data object (d)

Let $D_t(m)$ be the set of all data stored at a mobile node m at time t. Define $D_t^{wt}(m)$ as data object waiting to be sent by node m at time t and $D_t^{rc}(m)$ as data objects that have been received by node m before time t hence $D_t(m) = D_t^{wt}(m) \cup D_t^{rc}(m)$. Let $D_t(r)$ be the set of all data stored at a repository node r at time t and $D_t(r)$=$D_t^{wt}(r)$ where $D_t^{wt}(r)$ are data object waiting to be send by node r at time t.

To keep buffer size low, RDD removes any data object that reaches their expired time. At time t, the expiration time of d (denoted $x(d)$) is defined in Equation (4.1).

$$x(d) = ct(d) + ttl(d) \tag{4.1}$$

Note that the buffer may contain more than one data object for a given subscription channel. As there is a limit on how many messages can be passed in each interaction, finding the right prioritization for pushing these data objects is a crucial element. In this design, the data objects are ranked so that those due to expire sooner have a higher priority than others. At time t , the priority $p(d)$ of data object d is defined as

$$p(d) = \begin{cases} \frac{1}{x(d)-t} & \text{if} \quad \text{x(d)} > 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

Data objects are organized into different categories, a conventional way to access required data. A similar approach can be seen in [46][51] and a simpler form of such classification has been used in local advertisement websites, where the ads are classified into different channels. RDD organizes the data objects the protocol is intended to disseminate into specified channels according to content. We assume, that there are agreed or standard classification methods that used by all mobile to classify their data objects.

In our design, let C be a set channels, where $C = \{c_1, c_2, c_3, ..., c_{n_c}\}$. Each created data object d is assigned to exactly one channel on creation time and this channel belongs to the set C, i.e. $c(d) \in C$ and $|c(d)| = 1$.

4.3.2 User preference management

Each mobile node m subscribes to a set of channels $S(m)$ where $S(m) \subseteq C$. The subscription process is assumed to be carried out manually by the device users according to user needs although this could be automated based on previous history of content consumption. Therefore in a real system, this assumption could be easily relaxed in practical implementation.

4.3.3 Dissemination Protocol

The dissemination protocol of RDD is responsible for pushing data objects to intended mobile nodes. RDD is capable of delivering data objects both directly between MNs (i.e. without storing them in the buffers of relay nodes) and indirectly (i.e. by means of a store and forward technique). If both the source and an interested node encounter each other, the message is delivered using an underlying direct data dissemination protocol. Otherwise it is delivered using an underlying indirect data dissemination protocol, which is responsible for forwarding the messages to a repository, ideally with a high chance of successful delivery.

At the beginning of an encounter, a new connection is established if the two nodes are both free from any ongoing communication. During a connection, each node acts as both a data pusher and data receiver, for itself (mobile node only case) or on behalf of other nodes (repository node case). A connection is completed when both nodes have finished receiving data objects in which they are interested. In many cases a connection can terminate before that, simply because the two nodes have moved out of range or one of them has run out of battery power. When a new connection is established, the two nodes first exchange their subscribed channel set (S) , i.e. they tell each other what channels they are currently subscribed to.

4.3.3.1 Direct data dissemination algorithm

When two mobile nodes m_i and m_j encounter each other, they exchange their channels subscription sets $S(m_i)$ and $S(m_i)$. This exchange of channel subscription sets enables nodes to advertise channels they are subscribed in and this information is used to determine which data objects should be pushed for other node. The pseudo code of direct data dissemination algorithms is shown in Algorithm 4.1. Figure 4.3 shows diagram illustrates the communication process between two mobile nodes (m_i and m_j).

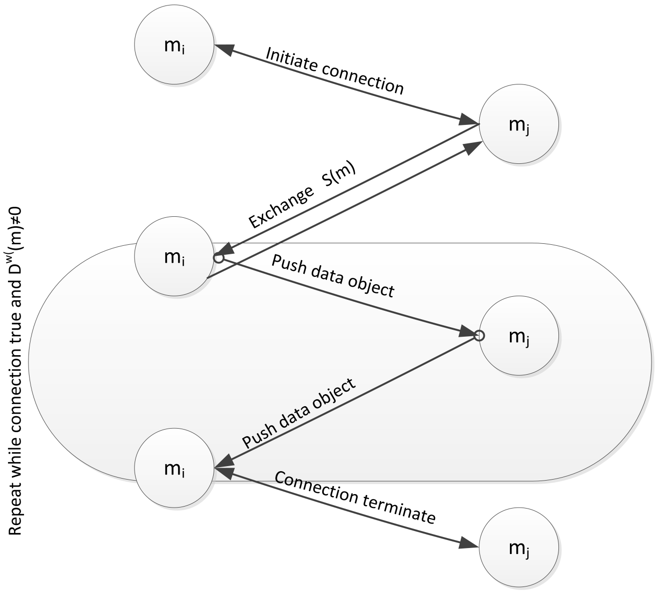


Figure 4.3: Diagram shows the communication process between mobile nodes

```
Initialization connection;
t=Current time;
Send $S(m_i)$ from $m_i$ to $m_j$);
Send ($S(m_j)$ from $m_j$ to $m_i$);
Let the list A be a copy of the elements of $|D_t^{wt}(m_i)|$ ordered by decreasing expire time ;
Let the list B be a copy of the elements of $|D_t^{wt}(m_j)|$ ordered by decreasing expire time;
while Connection true do
    $A_{rec}$=false; $B_{rec}$=false;
    while Not $A_{rec}$ and $|A|$ >0 do
        d = Pop(A);
        if x(d) $<$ t then
            if c(d) $\in$ S($m_j$) then
                if $d \notin D_t^{rc}(m_j)$) then
                    Push $d$ from $m_i$ to $m_j$; $D_t^{rc}(m_j) = D_t^{rc}(m_j) \cup \{d\}$; $A_{rec}$=True;
                end
            end
        else
            $D_t^{wt}(m_i)$ = $D_t^{wt}(m_i) - \{d\}$; $A = A - \{d\}$
        end
    end
    while Not $B_{rec}$ and $|B|$ >0 do
        d = Pop(B);
        if x(d) $<$ t then
            if c(d) $\in$ S($m_i$) then
                if $d \notin D_t^{rc}(m_i)$) then
                    Push $d$ from $m_j$ to $m_i$; $D_t^{rc}(m_i) = D_t^{rc}(m_i) \cup \{d\}$; $B_{rec}$=True;
                end
            end
        else
            $D_t^{wt}(m_j)$ = $D_t^{wt}(m_j) - \{d\}$; $B = B - \{d\}$
        end
    end
end
```

Algorithm 4.1: Direct data dissemination algorithm

4.3.3.2 Indirect data dissemination algorithm

As we mentioned above the repository is defined as a standalone fixed wireless device with no access to fixed infrastructure. Under RDD, the repository completely controls the multihop communication; therefore, the repository characteristics and its placement strategies used in RDD are introduced in section 4.5.

When a repository r encounters a mobile node m, the repository initiates the communication and m pushes their channel subscription $S(m)$. The repository uses $S(m)$ to determine which data objects from it's buffer will be pushed to the mobile node m. At same time the mobile node uses this interaction to push any data object in it's waiting buffer to the repository to be pushed to other subscribed mobile nodes when it encounteres. The pseudo code of indirect data dissemination algorithms is shown in algorithm

4.2.

```
Initialization connection;
t=Current time;
Send $S(m)$ from $m$ to $r$;
Let the list A be a copy of the elements of $|D_t^{wt}(m)|$ ordered by decreasing expire time ;
Let the list B be a copy of the elements of $|D_t^{wt}(r)|$ ordered by decreasing expire time;
while Connection true do
    $A_{rec}$=false;
    $B_{rec}$=false;
    while Not $A_{rec}$ and $|A|$ >0 do
        d = Pop(A);
        if x(d) < t then
            if $d \notin D_t^{wt}(r)$) then
                Push $d$ from $m$ to $r$; $D_t^{wt}(r) = D_t^{wt}(r) \cup \{d\}$; $A_{rec}$=True;
            end
        else
            $D_t^{wt}(m)$ = $D_t^{wt}(m) - \{d\}$ ; $A = A - \{d\}$
        end
    end
    while Not $B_{rec}$ and $|B|$ >0 do
        d = Pop(B);
        if x(d) < t then
            if c(d) $\in$ S($m$) then
                if $d \notin D_t^{rc}(m)$) then
                    Push $d$ from $r$ to $m$; $D_t^{rc}(m) = D_t^{rc}(m) \cup \{d\}$; $B_{rec}$=True;
                end
            end
        else
            $D_t^{wt}(r)$ = $D_t^{wt}(r) - \{d\}$ ; $B = B - \{d\}$
        end
    end
end
```

Algorithm 4.2: Indirect data dissemination algorithm

4.3.4 Repository and placement strategies

Placing the repository in the most suitable place is a challenging research problem, and will depend on knowledge of the mobile nodes movement and communication patterns. Under RDD, the repository is the only node responsible for message relay, so it completely controls the multi-hop message delivery. Therefore, the repository controls all indirect communications between MNs. As the repository location changes, the number of MNs with the opportunity to contact the repository may change. An increase in these contact opportunities should lead to an increase in message relay rate. Consequently, the repository performance strongly depends on its location, so optimizing the location for the repository is very important in the protocol design. As stated in RDD design assumptions, a data dissemination protocol is aimed at a campus sized environment, so the repository location problem is researched in an environment where human movement occurs in paths that are clearly defined, such as roads in a city map as seen in Figure 4.4. Repository can be be placed in road junctions shown in Figure 4.4.

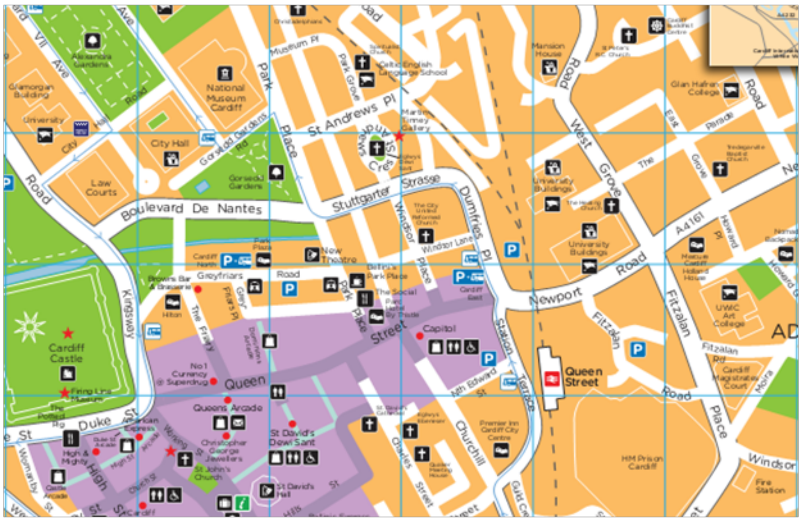


Figure 4.4: Roads can be modelled by junctions and edges, with off-locations representing venues.

In the last few years, we have seen emerging location-based social networking such as

Foursquare [20], GyPSii [53] , and Gowalla [24], which provide many services such as finding your nearby friends and letting your friends know, you are “checking in” to a location. While the Gowalla service has shutdown recently, Foursquare and GyPSii are still providing services. These applications employing mobility information of the users that mobile cellular networks can provide. The same mobility information can be easily used to generate traffic predictions according to historical patterns and can also easily be used to find the most crowded road or location, which is likely to be a suitable location for repository hence the most crowded locations usually the most suitable location to place a repository. However, in RDD two strategies “Betweenness” (B) and “Social Betweenness” (SB) are proposed to address the repository location problem. These two strategy are discussed in the following subsections.

4.3.4.1 Betweenness

The fact that the map is a graph means that the centrality of any vertex in that graph may be calculated by using the betweenness measure of how important a vertex is by counting the number of shortest paths that it is a part of. Vertices that occur on many shortest paths between other vertices have higher betweenness than those that do not [47]. Since humans usually choose the shortest path between their source and destination, the vertex with the highest betweenness in a graph should be expected to be a good place for the repository in that graph, simply because a lot of people pass that vertex. Let G be the graph, and let V be the set of vertices in G. The betweenness $B(v)$ for a vertex v is defined in Equation (4.3).

$$B(v) = \sum_{\substack{v_s \neq v \neq v_t \in V \\ v_s \neq v_t}} \frac{\sigma_{v_s v_t}(v)}{\sigma_{v_s v_t}} \tag{4.3}$$

where $\sigma_{v_s v_t}$ is the number of shortest paths from v_s to v_t, and $\sigma_{v_s v_t}(v)$ is the number of shortest paths from v_s to v_t that pass through a vertex v.

4.3.4.2 Social Betweenness

While the betweenness strategy is able to determine the highest centrality ranking vertex, this ranking is based on the assumption that MNs have equal preferences for all locations, which is not representative of real life Therefore, despite a vertex being considered as a highest betweenness vertex in the graph, it may not be passed often by mobile nodes because it is on a route to some locations not preferred by mobile nodes. An alternative strategy is needed to determine the repository location. In real life, users usually choose their destination according to their spatial relationship.

The preferences SpR for particular destinations defined in section 3.4 can be taken into account by extending *Betweenness* to give the new measure proposed, referred to as *Social Betweenness* (SB) measure. *Social betweenness measures how likely a vertex could be used by mobile nodes, based on the knowledge of the users preferences as well as the location of vertex.* The SB centrality of a vertex v in graph G is calculated from the probability of use of the shortest paths between each pair of vertices of the graph G passing by vertex v. Equation (4.4) defines the SB for vertex v.

$$SB\,(v) = \sum_{v_t \in V} \sum_{m \in M} \frac{\sigma_{h(m)v_t}(v).SpR(m, v_t)}{\sigma_{h(m)v_t}} \tag{4.4}$$

where $\sigma_{h(m)v_t}$ is the number of shortest paths from $h(m)$ to v_t, and $\sigma_{h(m)v_t}(v)$ is the number of shortest paths from $h(m)$ to v_t that pass through a vertex v.

The recent emergence of location-based social networking services such as Foursquare allows access to a rich store of information on user mobility and behaviour without destroying their privacy. Such information could be used as the input to derive *SpR* values.

Example 4.1: Let $G(V, E)$ be a graph with V representing locations and Junctions, and edges E representing the direct routes between these vertices. A vertex v define as junction if the v connection degree ≥ 2 while a vertex v define as location if the v

connection degree $= 1$. Figure 4.5 shows G with 30 vertices connected by edges. The distance between vertices are shown in the graph.

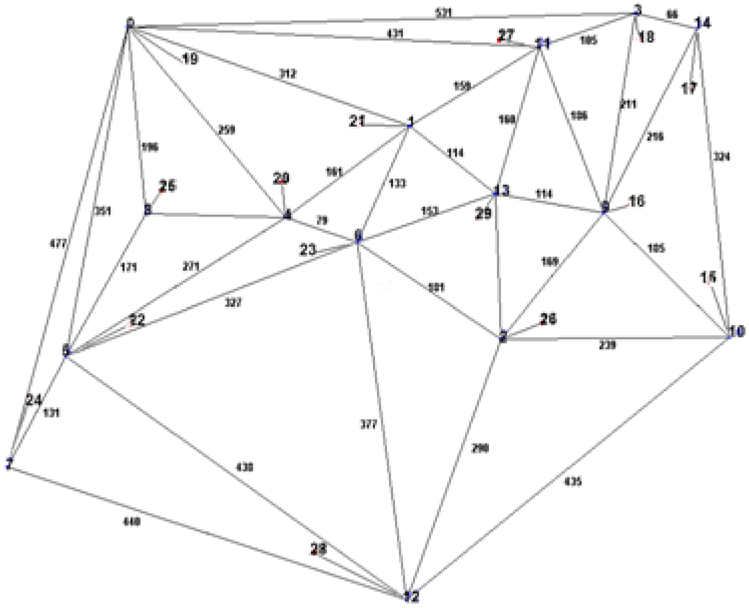


Figure 4.5: A Graph G with 30 vertices

The betweenness centrality counts how many shortest paths between each pair of vertices of the graph pass by a given node. Table 4.3 shows the betweenness value for each vertex calculated using Equation 4.1. Please note in our model, only betweenness of junctions is important as repositories will not be placed at off-location.

In Figure 4.5, the vertices numbered v_0 to v_{14} are junctions, i.e. $v_i \in J$, $i = 0, ..., 14$ while the vertices numbered v_{15} to v_{29} are considered are off-locations, i.e. $v_i \in F$, $i = 15, ..., 29$. Assume that each location is considered to be a home location for one mobile node only, i.e. there are 15 off-locations assigned to one of 15 mobile nodes (m_0 to m_{14}) . Each mobile node has an SpR value for each location, presented in Table 4.2, generated by MdNS using a Zipf distribution as discussed in chapter 3 with $\alpha = 1$. Note SpR has the value 0 for home locations.

	15	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**
m_0	0.00	0.03	0.04	0.05	0.10	0.06	0.15	0.03	0.02	0.02	0.03	0.08	0.03	0.31	0.04
m_1	0.15	0.00	0.06	0.04	0.02	0.02	0.04	0.03	0.31	0.05	0.03	0.03	0.08	0.10	0.03
m_2	0.31	0.04	0.00	0.15	0.10	0.03	0.06	0.03	0.02	0.08	0.05	0.03	0.04	0.02	0.03
m_3	0.03	0.06	0.04	0.00	0.15	0.08	0.31	0.02	0.02	0.05	0.03	0.10	0.03	0.04	0.03
m_4	0.10	0.15	0.03	0.08	0.00	0.03	0.31	0.03	0.03	0.05	0.06	0.02	0.02	0.04	0.04
m_5	0.15	0.03	0.08	0.02	0.31	0.00	0.03	0.02	0.06	0.04	0.04	0.03	0.03	0.05	0.10
m_6	0.04	0.06	0.02	0.10	0.03	0.15	0.00	0.08	0.05	0.03	0.04	0.31	0.02	0.03	0.03
m_7	0.03	0.02	0.03	0.06	0.04	0.03	0.10	0.00	0.03	0.02	0.04	0.15	0.08	0.05	0.31
m_8	0.04	0.04	0.31	0.08	0.10	0.03	0.02	0.06	0.00	0.03	0.15	0.02	0.03	0.03	0.05
m_9	0.03	0.31	0.03	0.03	0.03	0.04	0.10	0.04	0.05	0.00	0.06	0.02	0.08	0.15	0.02
m_{10}	0.03	0.03	0.02	0.06	0.10	0.03	0.15	0.31	0.03	0.08	0.00	0.04	0.02	0.05	0.04
m_{11}	0.02	0.15	0.05	0.03	0.04	0.08	0.10	0.03	0.03	0.03	0.06	0.00	0.02	0.04	0.31
m_{12}	0.31	0.15	0.02	0.03	0.08	0.04	0.03	0.03	0.10	0.04	0.03	0.02	0.00	0.06	0.05
m_{13}	0.15	0.02	0.06	0.02	0.05	0.04	0.08	0.04	0.10	0.03	0.03	0.03	0.31	0.00	0.03
m_{14}	0.10	0.03	0.31	0.05	0.02	0.04	0.03	0.04	0.08	0.15	0.06	0.03	0.03	0.02	0.00

Table 4.2: The SpR values for the mobile nodes

Applying the SB equation using the spatial preferences of mobile nodes to other locations presented in Tables 4.2, gives the values in Table 4.3. It is noticeable that SB value are different from the B value as can be seen in Table 4.3. The vertex with the highest B ranking (Table 4.4) is not necessary also the highest SB ranking as we see in Table 4.5. Because SB measures how often this vertex used by mobile nodes in addition to it's centrality in the map, we believe SB will be an effective method to find the best location for the repository specially users' locations preferences data can easily obtained nowadays from the cellular networks companies without violation of their privacy.

Vertex id	**B value**	**SB value**
v_0	28	106.49
v_1	64	222.96
v_2	46	147.4
v_3	44	134.96
v_4	70	188.59
v_5	48	152.25
v_6	68	217.11
v_7	28	83.45
v_8	28	84.35
v_9	46	181.16
v_{10}	28	122.81
v_{11}	54	171.33
v_{12}	30	100.31
v_{13}	52	217.98
v_{14}	30	107.05

Table 4.3: Vertices B and SB Values

Ranking	**vertex id**
1	v_4
2	v_6
3	v_1
4	v_{11}
5	v_{13}
6	v_5
7	v_2
8	v_9
9	v_3
10	v_{12}
11	v_{14}
12	v_0
13	v_7
14	v_8
15	v_{10}

Table 4.4: Vertices B ranking

Ranking	**vertex id**
1	v_1
2	v_{13}
3	v_6
4	v_4
5	v_9
6	v_{11}
7	v_5
8	v_2
9	v_3
10	v_{10}
11	v_{14}
12	v_0
13	v_{12}
14	v_8
15	v_7

Table 4.5: Vertices SB Ranking

4.4 Summary

This chapter presented, in detail, the design of the proposed data dissemination system for opportunistic networks. From the literature review in chapter 2, we highlighted the limitations of existing approaches which were taken into account in our design. We have set up three main objectives our system aims to meet, which are do not rely on mobile node cooperation, to be able to employ the mobility characteristics of users, and to be a decentralized system. To simplify our design, we have presented a set of assumptions at the start of this chapter. Since most of the proposed approaches rely on mobile nodes cooperation, we have taken a different approach which employs standalone fixed wireless devices to relay data objects instead of using mobile nodes. The RDD design composed of four parts, which are data object management, user preferences management, data dissemination protocol, and the repository placement. Firstly, in data object management, we have defined how data objects created, what form of data objects the system can deal with, and how the system manages data objects within the device buffers. Secondly, in user preference management, we have defined how the system organizes user channel subscriptions. Thirdly, in data dissemination, we have defined the dissemination protocol which is responsible for pushing data objects between devices. Finally, we have defined the placement strategy which is used by the system to place the repository and for this purpose we have defined social betweenness, a placement strategy which employs the knowledge of the user's preferences as well as the geographical location in determining the suitability of a location.

Chapter 5

RDD Evaluation and Results

This chapter presents the analyses of results in order to qualify the benefits of using the RDD system for data dissemination in opportunistic networks. The study was carried out using the evaluation model presented in chapter 3. First, the system performance metrics are presented which were used in different tasks in the evaluation process. Next, details of how the benchmark datasets created are presented. Before presenting the results we justify the validity of the simulation experiments. The evaluation process starts by comparing the repository location strategies presented in chapter 4. Next, using the best location strategy, RDD is evaluated by comparing network data delivery rate and redundancy rate with that of the epidemic protocol in identical network scenarios. Finally a conclusion of the evaluation experiments is presented.

5.1 Performance Metrics

Many metrics are commonly used to measure the performance of network protocols. This section presents the metrics used in our evaluation process.

Let D_T be the set of data objects created in the entire duration of the simulation time (denoted as T), i.e. time period $(0, T)$, where $D_T = \{d_1, d_2, d_3, ..d_{N_d}\}$. Let $D_t(m)$ refer to the subset of D_T carried by a given mobile node m at time t, and let $D_t(r)$ refer to the same for a repository node r at time t. The data objects carried by a mobile node m or similarly repository are placed into subsets based on whether they are available to be forwarded to further nodes or not. For a node m the maximum number of data objects are allowed to be saved by the node is defined as the buffer size for m. Figure 5.1 shows an overview of data objects management within a node itself and when two nodes encounter.

Received List: The list $D_t^{rc}(m)$ contains all data objects which have been $received$ by a node m in the time period $[0, t]$. As shown in Figure 5.1, a data object d is added to $D_t^{rc}(m)$ only due to the occurrence of a *Send event.*

- *Send event* (event 11 and 12 in Figure 5.1) is responsible for pushing a data object from a node s (either a mobile node or a repository) to a node m which is interested in that data object, i.e. if s encounters m and there exists $d \in D_t^{wt}(s)$ (where $D_t^{wt}(s)$ includes all data objects carried by s at time t which are waiting to be sent to other nodes) such that $c(d) \in S(m)$, and d has not already seen by m,(i.e. $d \notin D_t^{rc}(m)$) then s pushes d to m and it is added to $D_t^{rc}(m)$. Every time a data object d is added to $D_t^{rc}(m)$, the data objects in $D_t^{rc}(m)$ are sorted by expiration time of data objects in ascending order, i.e. data objects with the earliest expiration times $x(d)$ are listed first.

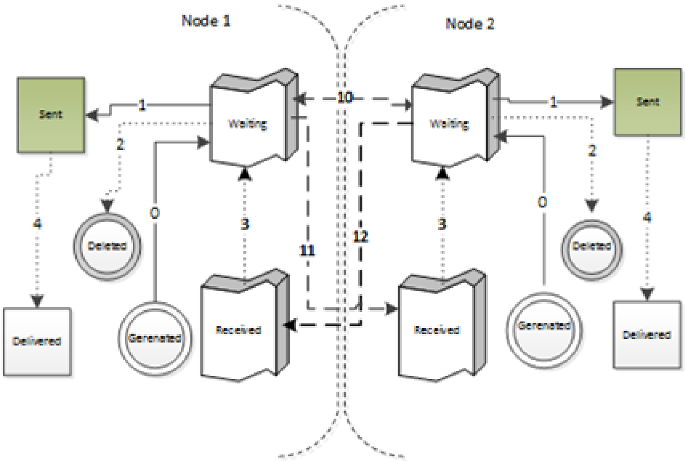


Figure 5.1: Data object status, showing movement between buffers and lists when two nodes encounter each other .

Waiting list : The list $D_t^{wt}(m)$ includes all data objects carried by m at time t which are waiting to be sent to other nodes. The list $D_t^{wt}(r)$ includes all data objects carried by r at time t which are waiting to be sent to other nodes. Note that any data objects $d \in D_t^{wt}(m)$ or $d \in D_t^{wt}(r)$are removed when $x(d) < t$.

As shown in Figure 5.1, a data object d belongs to the list $D_t^{wt}(m)$ for mobile node m due to the occurrence of one of the following:

- *A generating event* (event 0 in Figure 5.1) is responsible for generating a new data object by the node. I.e. if m generates a new data object d with channel $c(d)$ and time to live $ttl(d)$ at time t, (i.e. $ct(d) = t$), then d is added to $D_t^{wt}(m)$and $x(d)$ is defined in Equation (4.1).
- *Receive and relay event* (event 3 in Figure 5.1) follows a send event and is responsible for copying a data object from a node's received list to it's waiting list provided the node willing to relay. That is if m has $d \in D_t^{rc}(m)$ and m is willing to relay then d is copied to $D_t^{wt}(m)$ provided $x(d) < t$.
- *Relay only event* (event 10 in Figure 5.1) and is responsible for pushing a data object d from a node m_A to another node m_B which is not interested in the data object (i.e. $c(d) \notin S(m_B)$) but is willing to provide relay. That is if m_A has $d \in D_t^{wt}(m_A)$ and encounters m_B who is willing to relay and $d \notin D_t^{wt}(m_B)$ then m_A push d to m_B and d is added to $D_t^{wt}(m_B)$.

A data object d belongs to the list $D_t^{wt}(r)$ for a repository r only due to the occurrence of the *Relay only event*:

- *Relay only event* (event 10 in Figure 5.1) is responsible for pushing a data object d from node m to another node. I.e. if m has $d \in D_t^{wt}(m)$ and encounters r and $d \notin D_t^{wt}(r)$then m push d to r and d is added to $D_t^{wt}(r)$.

Every time a data object d is added to $D_t^{wt}(m)$, the data objects in $D_t^{wt}(m)$ sorted by expiration time of data objects in ascending order, i.e. a data object with earliest expiration time $x(d)$ listed first. Although we have assumed that buffer size is unlimited, any data objects $d \in D_t^{wt}(m)$ or $d \in D_t^{wt}(r)$ are removed from the buffer when $x(d) < t$.

In addition to these two lists, there are other two counters (*sent* and *delivered*) which have been defined for analysis purposes.

1. The ***Sent*** counter, counts all data objects that have been sent by a node (mobile or repository) using (relay-only and send events) during the simulation regardless of the subscription of their destinations. $C_T^s(m)$ is a counter for the number of data objects has been sent by a node m regardless of their destinations and $C_T^s(r)$ is a counter for the number of data objects has been sent by a repository r regardless of their destinations.

1. The ***Delivered*** counter, which counts all data objects that have been sent successfully to a target, i.e. $C_T^d(m)$ is a total number of all data objects received by m where $c(d) \in S(m)$.

The metrics used in our evaluation process are presented now, while the following metrics are (technically) only defined for RDD, but used for EPR too. These metrics are as follows:-

1. Data Delivered (DD): This metric gives an indication of the ability of the protocol to deliver data to its destinations, i.e. how many messages successfully delivered to interested nodes. DD is defined as the total number of messages have been successfully delivered to its destinations for all mobile nodes in the entire duration of the simulation time and is defined by Equation (5.2).

$$DD = \sum_{m \in M}^{|M|} C_T^d(m) \tag{5.1}$$

2. Data Sent (DS): this metric gives an indication of the traffic generated by the protocol and is the total number of messages sent by all nodes in the entire duration of the simulation time and is defined by Equation (5.3).

$$DS = \sum_{m \in M}^{|M|} C_T^s(m) + \sum_{r \in R}^{|R|} C_T^s(r) \tag{5.2}$$

3. Data Redundancy (DR): this metric gives an indication of the number of messages sent without immediate benefit for final delivery, i.e. the traffic overhead created by the protocol. DR is the total number of data objects sent by all mobile nodes minus the number of data objects delivered by all nodes and is defined by Equation (5.4).

$$DR = DS - DD \tag{5.3}$$

5.2 Benchmark Datasets

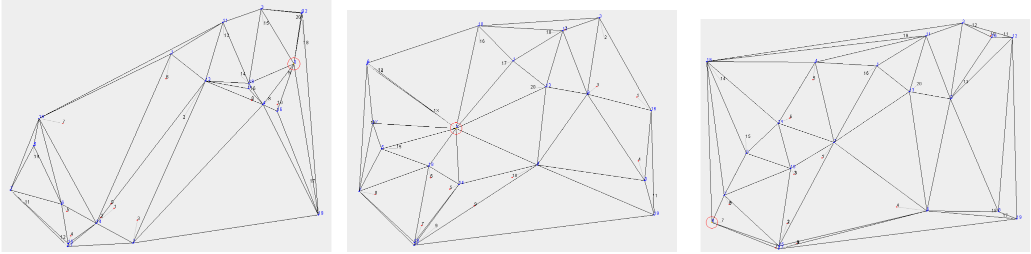
In order to model a scenario for evaluation in the opportunistic network, we define four sets of information: *environment model*, *users' mobility patterns*, *data objects generation*, and *users' preferences*. Since there are no available data source offering complete sets of such information, it has been necessary to create artificial, yet realistic benchmark datasets. In this section, we describe in detail, how these benchmark datasets have been created.

Many of the simulation programs make use of a pseudo-random number generator with a uniform distribution to create a random value in specific range. We have used a pseudo-random number formula depends on two random seeds to create a sequence of random numbers in specific range. During our simulation experiments, we have used 10 sets of seeds, 7 sets of seeds used for creation of 7 mobility patterns and data generation and 3 used for the generation of three distinct graphs. In section 5.3, we have carried out a statistical analysis that shows using this limited number of seeds will be sufficient to draw statistically significant conclusions from our evaluations.

5.2.1 Environment model

In our mobility model, a graph is used to mimic a real environment map. As discussed in chapter 3, a graph is created by distributing the chosen number of vertices (V) randomly in the fixed size simulation area and connecting them by their Delaunay triangulation. In the created graph, V is composed of two disjoint, non-empty subsets of vertices named *Junctions* (J) and *Off-Locations* (F). The number of edges connected to vertex v is defined as the degree of a vertex, $g(v)$. Recall that in the generated graph $g(v) = 1\,\forall v \in F$ and $g(v) \geq 2\,\forall v \in J$. We define the number of $v \in F$ which are allowed to attach to vertex $v \in J$ as the attach degree, $H(v)$. In the generated graph $H(v) \geq 1\,\forall v \in J$. Without loss of generality the default value is that $H(v) = 1\,\forall v \in J$, i.e. $|F| = |J|$, $H(v)$ is configurable as we will see in some experiments within this chapter.

In our evaluation process for each test we have used 3 different graphs generated using three different sets of seeds. Figure 5.2 (a, b, and c) shows an example of three graphs generated with $|J| = 20$ vertices and $H(v) = 1\,\forall v \in J$, i.e. $|J| = |F| = 20$.



(a) Map 1 (b) Map 2 (c) Map 3

Figure 5.2: Three distinct maps with the same number of vertices

5.2.2 Mobility patterns

The simulation of user mobility patterns is modelled using HMM as presented in chapter 3. There are two main factors which play an important role in the mobile nodes mobility patterns, namely, off-time duration and destination selection. In our simulation, to generate a different mobility pattern for each mobile node, a separate random seed is used to control the off-time and destination selection. In the following paragraphs, we will discuss the off-time and destination selection.

- **Off-time**: As we have described in the mobility model, each mobile node stays in at least in one available *off-location* ($v \in F$) for a duration of time, this time defined as *off-time*. Hence every time a mobile node m reaches an *off-location*, m randomly chooses a new off-time from the defined range. In our simulation setting, we provide the minimum and the maximum values for off-time, i.e off-time=$[t_{min}, t_{max}]$. For each simulation run a new set of off-times for the same mobile node is created by using a new random seed.
- **Destination selection:** For each mobile node, we have defined spatial relationship using Zipf's power law with $\alpha = 1$ as described in chapter 3. Mobile node m will choose a destination $f \in F$ according to their SpR values using roulette wheel selection. For each simulation run, we use a new random seed to produce a different roulette wheel selection and consequently a different destination selection for the same mobile node.

5.2.3 Data objects generation

In our simulation, we have assumed that each mobile node creates and injects a number of data objects into the network during the simulation time. For the number of data objects created by each mobile node, we assume that an average of number of data objects created by each mobile node is provided by the simulation setting. In our simulation, we have created data objects according to a Poisson process [34] with rate λ which is equal to the average of number of message to be created by a mobile node during the first half of simulation time interval, i.e. time interval equal to [0,T/2] where T is entire duration of the simulation time . The main point of creating the all messages in the first half is to allow sufficient time for the messages to have a reasonable chance of delivery within the simulation time. To allow the reuse of identical message creation patterns across scenarios, a separate random seed is used from that used for mobility patterns.

Two important aspects of created data objects are the data object channel category and expiration time which are assigned at creation time. As we described in Chapter 4, each data object d is categorized by a specific channel, $c(d)$. In our simulation this is uniformly randomly assigned at the creation time. We also provide the minimum and maximum values for time to life (ttl) range for data objects. For each created data object d we randomly assign $ttl(d)$ within the provided range.

5.2.4 User preferences

In our simulation, two main user preferences were taken into account. The first is the users willingness to relay data objects on behalf of other nodes which we treat as a binary decision. For each mobile node m, $W(m)$ takes the value 0 if the node is not willing to relay and $W(m)$ takes the value 1 if it is willing to relay.

Secondly, we define user channel subscription. In our simulation, we have defined 10 different channels, i.e. $C = \{c_1, c_2, ..., c_{10}\}$. Each data object d is assigned randomly to one of these channels upon its creation. At initial experiments, we assume that mobile node m can be either subscribed to all channel, i.e $S(m) = C$ or none, i.e. $S(m) = \phi$.

5.3 Validity of experiment runs

In our evaluation process for each test, the presented results are an average of 21 simulation runs over 3 different (with equal number of vertices) maps and 7 different mobility patterns for each map for each test). Before exploring the performance of RDD, we demonstrate that using 3 maps and 7 mobility scenarios are sufficient to draw statistically significant conclusions.

To justify using a limited number of seeds, a statistical analysis was carried out using ANOVA Two-Factor without Replication [25]. DD was used as the main metric in this analysis. The ANOVA results shows that the differences in obtained values using 3 different maps and 7 mobility patterns are not statistically significant. From this test, we have concluded that using this limited number of seeds will be enough to draw up a conclusion for our evaluation process.

Throughout our evaluation process, we have conducted 21 simulation runs for each test, which requires a large number of simulation runs in total. To run this number of simulation tests, we used the distributed resources available at the Cardiff University Condor Pool. Condor (which became HTCondor in October 2012) is a high throughput computing system developed by the Condor Team at the University of Wisconsin in 1988 [61]. It enables users to distribute a computational task across hundred of PCs. The extra processing power allows large tasks to be completed quicker. The Condor Pool consists of a central manager, submit machines and compute nodes. When a user submits jobs, the condor central manger distribute them between condor compute nodes and when the jobs finish the manger return the jobs and any results to the submit machine. As an indication, using a windows 64 bit machine with Intel Core i3 2.5 GHz processor and 4 GB RAM, running one simulation test (single seed) for the repository location experiment with $|J| = 25$ and $|M| = 25$ takes approximately 10 minutes, giving a total times of 210 minutes. However using condor the 21 tests can be performed in about 30 minutes.

5.4 Simulation Parameters

The simulation parameters are an important factor in the evaluation process. In the first part of this section, we describe the simulation parameters that have been used in our evaluation process while in the second part we present and justify the default simulation values.

5.4.1 Simulation parameters

This section describes the simulation parameters that have been used in our evaluation process.

1. **Simulation time**: How many time steps the simulation experiments will run.
2. **Node communication range**: The communication range for nodes in meters.
3. **Node speed**: The distance a node travels in each time step (provided it is not in an off-time period).
4. **Average number of messages created**: The average number of messages created by each mobile node during the simulation time.
5. **Environment size**: The size of simulated environment as x-y dimension.
6. **Minimum & Maximum Off-time**: The minimum and maximum off-time for MNs.
7. **Number of data channels**: This defines how many channels of content have been defined.
8. **Number of messages passed per encounter**: The number of messages that can pass in one direction from one node to another in each encounter. This parameter is mainly used to mimic real life, where a moving node usually has very

limited time to communicate, and consequently a limited number of messages can transfer each time.

9. **Data handling priority**: The way data object sorted in data objects buffers see section 5.1.
10. **Minimum and maximum** $ttl(d)$: The minimum and the maximum time to life for the data objects.
11. **Number of mobile nodes**: The number of mobile nodes used in the simulation experiment, i.e size of M set or |M|.
12. **Number of Repository**: The number of repository used in the simulation experiment, i.e size of R set or |R|.
13. **Number of vertices**: The number of vertices $v \in J$ used to create the graph.
14. **Home locations density**: As we have discussed in chapter 3, $H(v) \geq 1 \forall v \in J$ where H(v) is number of $v \in F$ allow to attach to each $v \in J$, i.e. if $H(v) = 1 \forall v \in J$ then $|F| = |J|$.
15. **Density of MNs with** $W(m) = 1$**:** The percentage of MNs with $W(m) = 1$ within the simulated MNs
16. **Density of MNs with** $S(m) = C$**:** The percentage of MNs with $S(m) = C$ within the simulated MNs

5.4.2 Default value for the simulation parameters

Table 5.1 presents the default values for the simulations. These are used in all experiments unless stated otherwise.

Parameter	**Value**
Simulation Steps	*86400 steps* will be discussed in Section 5.5*
Communication range	*2 meters*
Speed	*1 meter/step*
Average messages created	*3 messages*
Environment size	*1000 * 700 meters*
Minimum off Time	*7200 steps* will be discussed in Section 5.5*
Maximum off Time	*21600 steps*
Number of Channels	*10 Channels*
Messages allowed to pass	*2 messages*
Data handling priority	$x(d)$
$ttl(d)$ *minimum time*	*21600 steps*
$ttl(d)$ *maximum time*	*43200 steps*
Buffer Size	*Unlimited*
$\|M\|$ No. of mobile nodes	25
$\|R\|$ No. of repositories	1
$\|J\|$ No. of junctions	25
$H(v)$	1
Density of MNs with $W(m) = 1$	0%
Density of MNs with $S(m) = C$	100%

Table 5.1: Simulation Parameters

The default values presented in Table 5.1 are based on the following assumptions:-

1. Humans practice discontinuous daily walking activity in the first 12 hours of the day.
2. Human off-time duration between 1-3 hours.
3. Average number of messages created by each mobile node in 6 hours is 3 messages

5.5 Simulation time model

In this section, we present how the time is modelled in our simulation. The justification of defining the simulation time by 86400 steps and off-time between 2100 steps to 43200 steps are now discussed.

1. **Simulation Time** The duration of the simulation experiment should mimic the real environment. Because the real environment here means a mobility pattern of real people, understanding the time of the real environment enable a simulation time to be defined. The fact of the day composed of 24 hours, where human usually rest for 8 hours, leaving only 16 hours for daily activity, which may include walking. Of this, people use up to 10-12 hours in their outdoor walking activity. In the simulation, time is measured by the number of steps instead of clock time for two reasons. First, steps eliminates any effect of hardware specification on the result and secondly, steps makes it easier to calculate how to mimic real life scenarios. Human average walking speed is between 90 to 120 paces per minute [42]. For simplicity, we define the mobile node speed as 120 steps/minute. Therefore if an individual walks for 7200 paces, it means he walks for approximately one hour. In our simulation, the default simulation time equal to 86400 steps which equivalent to 12 hours.

2. **Off-Time** This parameter defines the duration of the time where mobile node is not moving. The value of off-time is changed every time the mobile node reaches an off-location, therefore in our simulation we have defined a minimum and maximum time for the off-time and used a uniform random distribution to calculate new off-time after each arrival at destination. During 12 hours of activity, humans usually have a lot of time where no walking activity occurs, which is defined as off time. While determining the duration of off-time for a mobile node is beyond the scope of this research, it is assumed that off-time duration would be between 1-3 hours.

5.6 Evaluation of Repository Location Strategies

In chapter 4, two repository location strategies were introduced: Betweenness (B) and Social-Betweenness (SB), which will be evaluated in this section. In the investigation, DD is used as the main metric to determine the performance of the proposed strategies.

5.6.1 Experimental scenarios

To study the performance of the proposed strategies, a series of simulation experiments were carried out. The following experiments examine the two repository location strategies (SB and B) in four simulation scenarios (Table 5.2). The simulation setting used the default value presented in Table 5.1.

-	**Scenario 1**	**Scenario 2**	**Scenario 3**	**Scenario 4**
$\|M\|$	10 nodes	10 nodes	25 nodes	25 nodes
$\|J\|$	20 vertices	50 vertices	50 vertices	100 vertices
$\|R\|$	1 node	1 node	1 node	1 node

Table 5.2: Simulation scenarios for the evaluation of R location strategies

The first scenario is considered as small and dense, the second is small and sparse, the third scenario is large and dense, and fourth is a large and sparse scenario. The evaluation process investigated the performance of each scenario using 3 different maps, and for each map the experiment has been run in 7 different mobility patterns. Therefore for each scenario, 21 experiments were carried out.

5.6.2 Test Results

This experiment examines the effect of the repository location on the DD. The repository is placed at each $v \in J$, and the DD evaluated in each time. Figures 5.3-5.6 shows the performance comparison between using B and SB strategies in four simulated scenarios. In Figures 5.3-5.6, the x-axis represent the repository location ranking according to B for the betweenness graph and according to SB for the social betweenness graph. The ranking start with highest (ranking 1 in x-axis) to lowest which equal to $|J|$ used in corresponding test

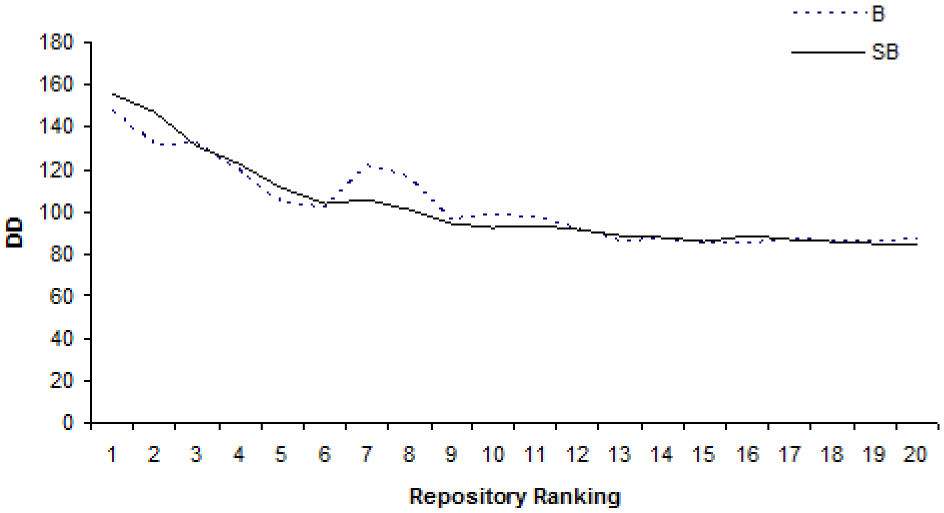


Figure 5.3: Comparison of B and SB performance in scenario 1

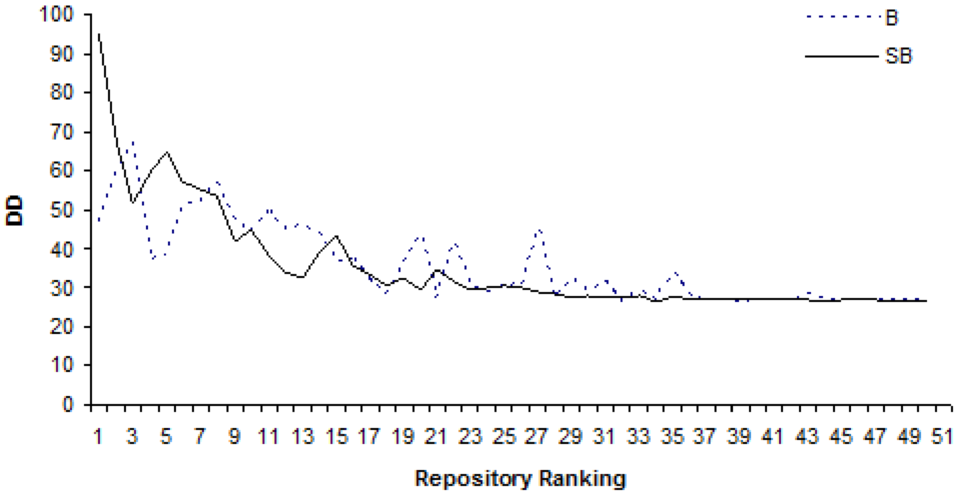


Figure 5.4: Comparison of B and SB performance in scenario 2

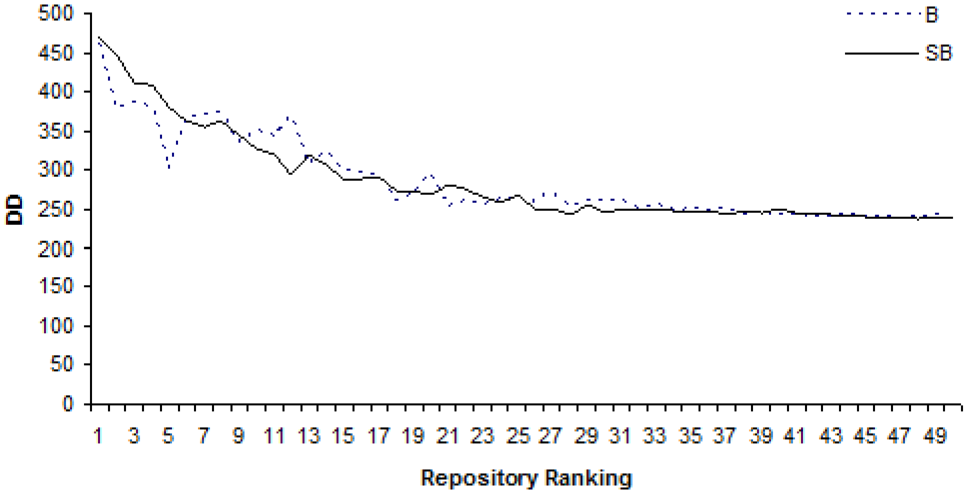


Figure 5.5: Comparison of B and SB performance in scenario 3

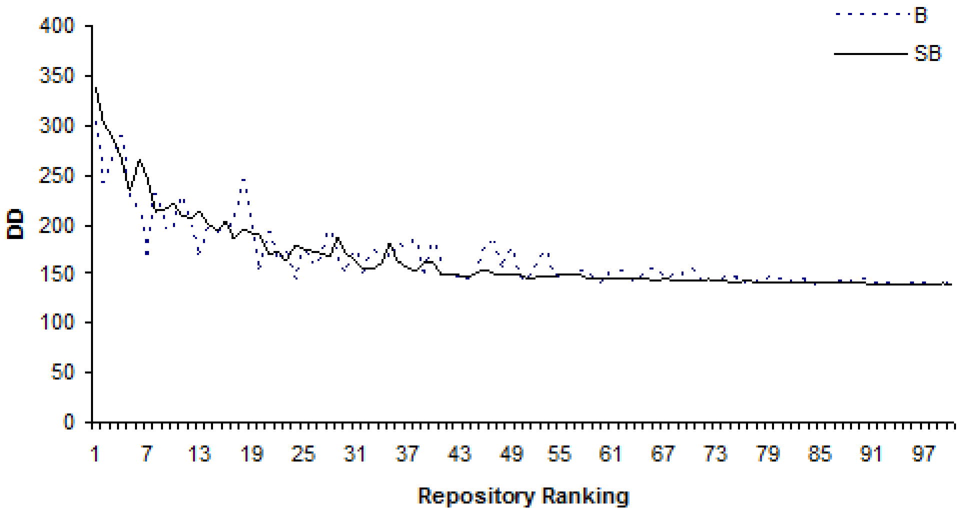


Figure 5.6: Comparison of B and SB performance in scenario 4

Despite the different simulated scenarios (1-4), Figures (5.3-5.6) shows that the performance of the DD often improves as the repository location B & SB value increases nevertheless we can notice using B as a location strategy, the DD sometime shows unpredictable change which represent the inconsistency of this strategy. In the four tested scenarios (1-4), the results show that using SB as a location strategy for the repository provides more consistent and smooth result compared to B, therefore we conclude that we should use SB as a repository location strategy and from now on all results presented in this thesis are based on a single repository located in the best SB location.

5.7 RDD Performance

To evaluate RDD, simulation experiments were performed comparing RDD with epidemic routing (EPR).

5.7.1 EPR as performance benchmark

The EPdemic Routing (EPR) is a popular protocol for data delivery in opportunistic networks (discussed in detail in section 2.5.1.1). The EPR relies on the flooding concept. The main idea of this protocol is as follows. If a carrier of a message encounters another node, it will deliver all the messages it carries that the other node does not already have. This approach can achieve the highest possible delivery ratio if it is provided with infinite bandwidth and buffer resources. Therefore it is considered as a good benchmark for comparison with alternative protocols.

Since EPR was designed as a forwarding protocol (data delivery to a specific destination), some modification to this protocol was necessary so that it could be used as a data dissemination protocol (where data is delivered to all nodes encountered). Here, EPR has been implemented with three settings to the hop limit. The hop limit for a message is the maximum number of hops that the message is allowed to travel to reach its destination. These settings are 2 hops limit (2h), 4 hops limit (4h) and unlimited hops (Uh).

5.7.2 Experiment setting

The simulation setting used the same default values presented in Table 5.1 while the protocol (RDD and EPR) settings are presented in Table 5.3.

Parameter	**RDD**	**EPR**
Hop limit	n/a	2,4, and unlimited hops
Repository location strategy	SB	n/a
$\|M\|$	25	25
$\|R\|$	1	0
$\|J\|$	25	25
$H(v)$	1	1
Density of MNs with $W(m) = 1$	0%	100%
Density of MNs with $S(m) = C$	100%	100%

Table 5.3: Experiment setting

Figure 5.7 shows that the EPR approach provides better DD despite the hop limit. Because the density of MNs with $S(m) = C$ is 100% there is no redundancy in EPR, while as we can see there is some redundancy associated with RDD. This occurs since messages routed through the repository require more than one hop, and some messages forwarded to a repository may never reach a final destination.

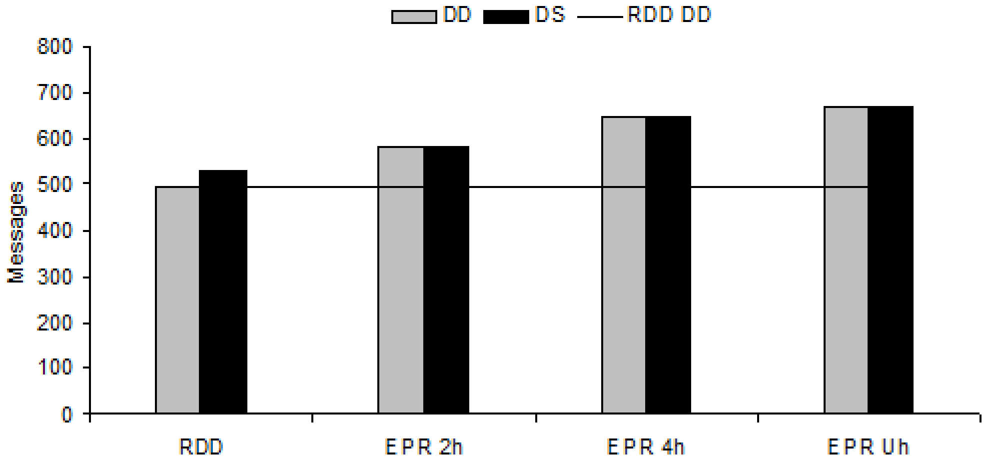


Figure 5.7: RDD vs. EPR performance

Further experiments have been carried out with varied densities of mobile nodes. We define the mobile node density as the number of mobile nodes $|M|$ divided by the number of junction vertices $|J|$ in the graph used the corresponding scenario. Table 5.4 presents the densities used in these experiments.

Scenario	1	2	3	4	5	6	7	8
$\|J\|$	25	25	25	50	50	50	50	50
$\|M\|$	12	25	50	15	25	35	40	50
$H(v)$	1	1	2	1	1	1	1	1
EPR hop limit	Unlimited							
Mobile node density	48%	100%	200%	30%	50%	70%	80%	100%

Table 5.4: Mobile density scenarios

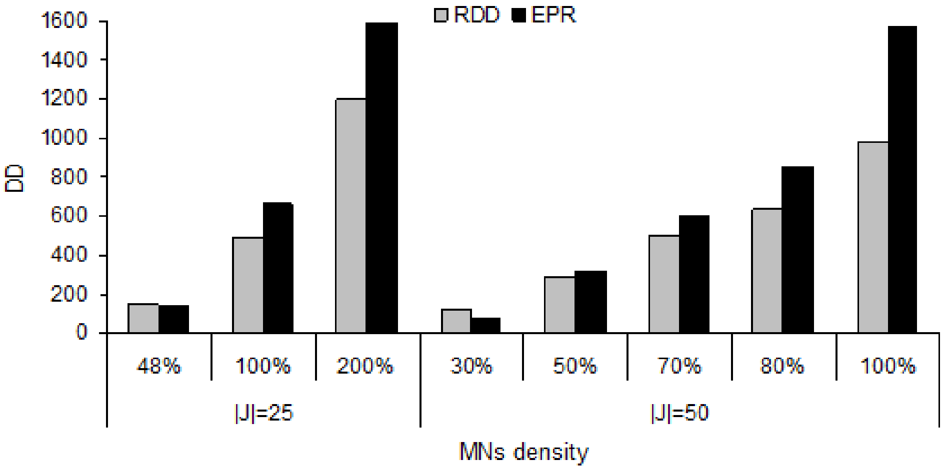


Figure 5.8: Impact of mobile nodes density on protocols performances

For a graph with $|J| = 25$, Figure 5.8 shows that when the density of mobile nodes equals 48% or lower, RDD performance is better than EPR, while when the mobile node density is 100% or 200% the performance of EPR is better. For graph with $|J| = 50$, the Figure 5.8 shows that when density of mobile nodes is equal 30%, the

RDD performance is the better than EPR, while when the mobile node density is 50% the EPR starts to perform better than RDD. From the above we can draw the conclusion that RDD will perform better than EPR in low density scenarios despite the hop limits, while EPR will perform better than RDD in middle or high density scenarios. This finding goes with fact that more mobile nodes means more relay for EPR so the EPR performed better in high density of mobile nodes.

5.8 RDD in partial cooperative environment

One of the arguments for using a repository in data dissemination is that the MNs are not usually willing to work as a relay or MNs are willing but it may be inefficient to work as relay. This test compares the performance of RDD against EPR in an environment where the density of MNs with $W(m) = 1$ is varied. Note that RDD requires no willingness, i.e. the repository is guaranteed willing to relay. The main point of this test is to investigate the performance of both protocols in an environment closed to real life environment where mobile node are not always willing to carrying messages to other nodes. Note that throughout our evaluation process, we represent the density of mobile nodes willing to cooperate as follows, “25% W” refers to the density of mobile nodes willing to cooperate being 25%.

5.8.1 Experiment setting

The simulation setting uses the same default values presented in Table 5.1 while the protocol (RDD and EPR) settings are presented in Table 5.5. For this test, we have defined four scenarios where the density of mobile nodes willing to relay varies as shown in Table 5.5. When this density is less than 100%, the selection of nodes that are willing to relay is made uniformly random.

Parameter	**RDD**	**EPR**
Hop limit	n/a	2,4, and unlimited hops
Repository location strategy	SB	n/a
Number of Repository	1	n/a
$\|M\|$	25	25
$\|R\|$	1	0
$\|J\|$	25	25
$H(v)$	1	1
Density of MNs with $W(m) = 1$	0%	25% (scenario 1)
-	-	50% (scenario 2)
-	-	75% (scenario 3)
-	-	100% (scenario 4)
Density of MNs with $S(m) = C$	100%	100%

Table 5.5: Experiment setting

Figure 5.9 shows that if the percentage of MNs that are willing to relay is less than less than 50%, RDD performs better despite the hop limits used by EPR. When the density reaches 75%, EPR performs better. The improvement EPR gains is up to 15% when 75% of mobile nodes are willing to cooperate and 17-36% when 100% of MNs willing to cooperate.

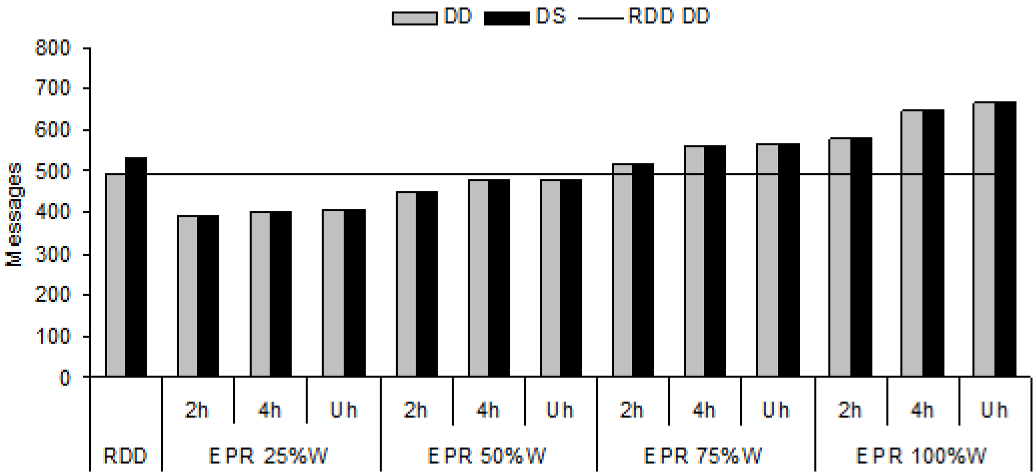


Figure 5.9: Impact of density of MNs with $W(m) = 1$ on the performances

5.9 RDD in partial subscription scenarios

In real life scenarios the mobile node channel subscription could have the following setting: $S(m) = \emptyset$ or $S(m) = C$ or $S(m) \subset C$. As we have stated above, at this stage we have assumed that mobile node m can be either subscribe to all channel, i.e. $S(m) = C$ or not subscribed, i.e. $S(m) = \emptyset$. This investigation studies the performance of RDD compared with EPR in different scenarios, where the density of subscribed mobile nodes varies. Note that throughout our evaluation process, we represent the density of subscribed mobile nodes as follows: "25% S" refers to the density of subscribed mobile nodes being 25%.

5.9.1 Experiment setting

The simulation used the same default values presented in Table 5.1, while the protocol (RDD and EPR) settings are presented in Table 5.6. For this test, we define three scenarios where the density of mobile nodes with $S(m) = C$ is set to 25%, 50%, and

75%. However, $W(m) = 1$ for all mobile nodes in these scenarios. The subset of MNs with $S(m) = C$ are selected uniformly randomly.

Parameter	*RDD*	*EPR*
Hop limit	n/a	2,4, and unlimited hops
Repository location strategy	SB	n/a
Number of Repository	1	n/a
$\|M\|$	25	25
$\|R\|$	1	0
$\|J\|$	25	25
$H(v)$	1	1
Density of MNs with $W(m) = 1$	0%	100%
Density of MNs with $S(m) = C$	25% (scenario 1)	25% (scenario 1)
-	50% (scenario 2)	50% (scenario 2)
-	75% (scenario 3)	75% (scenario 3)
-	100% (scenario 4)	100% (scenario 4)

Table 5.6: Tested subscription scenarios

Figure 5.10 shows that if the density of MNs with $S(M) = C$ is 25%, EPR provides the better DD rate. The improvement provided by EPR in DD is approximately 40% compared to RDD. However, this improvement comes with a high cost in DR with an average of 900% (over 2h, 4h, Uh scenarios) compared to RDD. When density is 50%, EPR still maintains higher DD with roughly the same percentage improvement, but with lower cost in DR with an average of 600% compare to RDD. When subscription density reaches 75% the DR cost getting lower to an average of 250% compared to RDD. Therefore, despite the improvement in DD provided by EPR, the high cost of DR makes RDD more practical and efficient to use. It is noticeable that the DR decreases in EPR when the percentage of MNs subscribed increases.

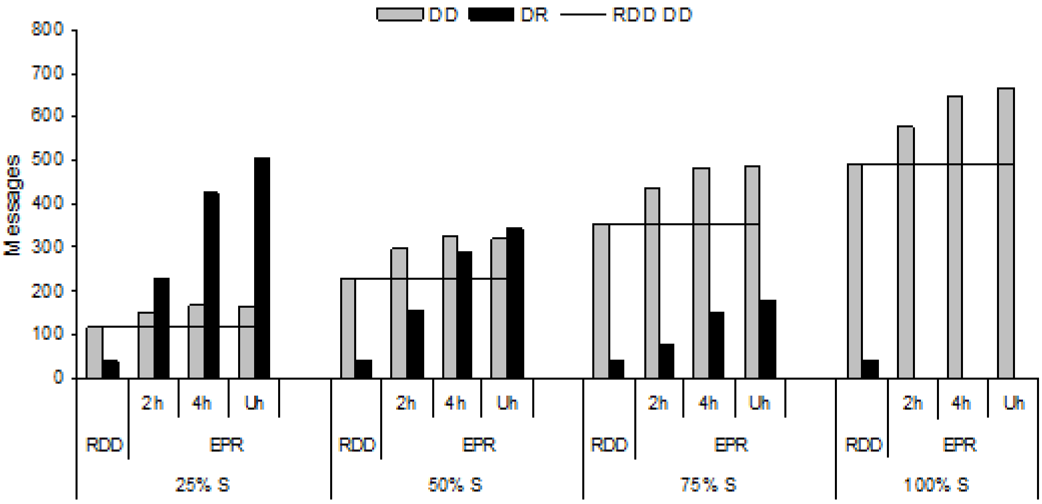


Figure 5.10: **Impact of density of MNs with S(m)=C**

The results in figure 5.10 show that in practical scenarios, where not all user require information, RDD is able to significantly reduce overheads, with only a small impact on delivery.

5.10 Subscription and cooperation density

This experimentation studies the performance of RDD compared with EPR in 9 different scenarios setting as presented in Table 5.7. The main point of the test is to evaluate the performance of both protocols in environments more close to reality where neither the density of MNs with $W(m) = 1$ nor the density of MNs with $S(m) = C$ are equal 100% but it has varied values. This is achieved by considering different levels of subscription (25,50,75%) and cooperation (25,50,75%) simultaneously.

5.10.1 Experiment setting

The simulation uses the same default values presented in Table 5.1 while the protocol (RDD and EPR) settings are same as those presented in Table 5.6. The tested scenarios are presented in Table 5.7.

Density of MNs with $W(m) = 1$	25%	25%	25%	50%	50%	50%	75%	75%	75%
Density of MNs with $S(m) = C$	25%	50%	75%	25%	50%	75%	25%	50%	75%

Table 5.7: Tested scenarios in different S and W density

Figure 5.11 shows the performance of the two protocols when the density of mobile nodes with $S(m) = C$ is 25%. The Figure shows that RDD produces better DD irrespective of the hop limit when the density of mobile nodes with $W(m) = 1$ is 25%. When the density of mobile nodes with $W(m) = 1$ is 50%, the RDD provides the same DD in EPR approximately but with a higher cost of DR with an average of 330% compared to the RDD. When the density of mobile nodes with $W(m) = 1$ is 75%, the EPR provides better DD compared to RDD but higher cost of DR with an average of 530% compared to the RDD.

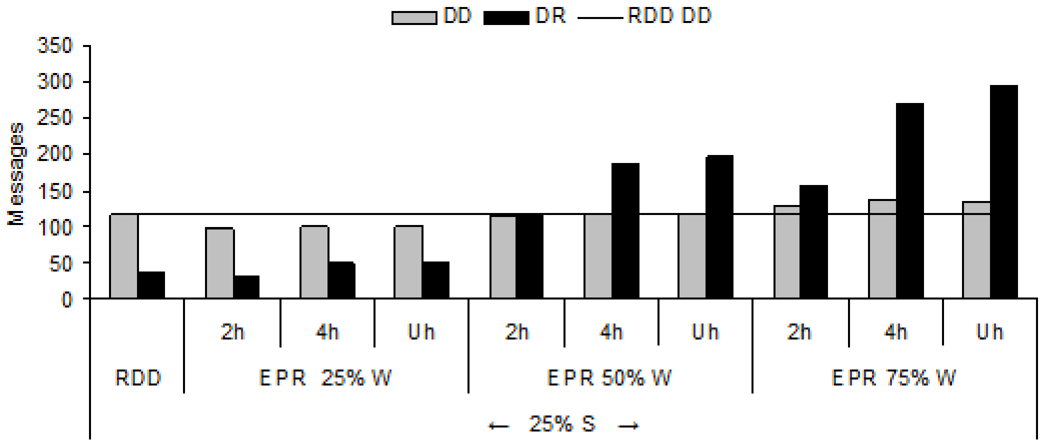


Figure 5.11: Performance in scenarios where MNs with $S(m) = C$ density is 25%.

Figures 5.12 shows the performance of the two protocols when the density of mobile nodes with $S(m) = C$ is 50%. The Figure shows that RDD provides a better DD when the density of MNs willing to cooperate is 25%. RDD is still able to provide better DD when the density of MNs willing to cooperate is equal to 50% and hop limit is less than 4 hops. When the density on MNs willing to cooperate is 50% and hop limit is greater than 4 hops, EPR provides a very small improvement. However, this improvement in DD comes with a high cost of DR with an average 150% compared to DR in RDD. When the density of MNs willing to cooperate is 75% EPR provides better DD but with the high cost of RD with an average of 240% compared to DR in RDD.

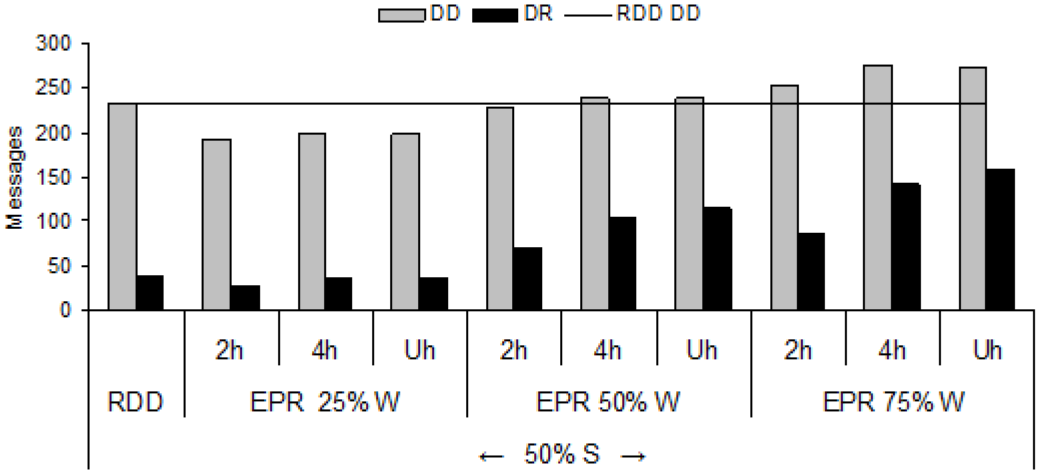


Figure 5.12: Performance in scenarios where MNs with $S(m) = C$ density is 50%.

Figure 5.13 shows the performance of the two protocols when the density of mobile nodes with $S(m) = C$ is 75% . The Figure shows that RDD provides better DD when the density of MNs willing to cooperate is 25%. RDD is still able to provide better DD when the density of MNs willing to cooperate is equal to 50% and hop limit is less than 4. When the density of MNs willing to cooperate is 50% and hop limit is greater than 4 hops the EPR and RDD provides equal DD. When the density of MNs willing to cooperate is 75% EPR able to provide an improvement in DD by 9% compared to RDD. This improvement comes with higher cost of DR with an average of 36%

compared to RDD. It's also noticeable that when the density of subscribed MNs is higher, EPR able to provide less DR.

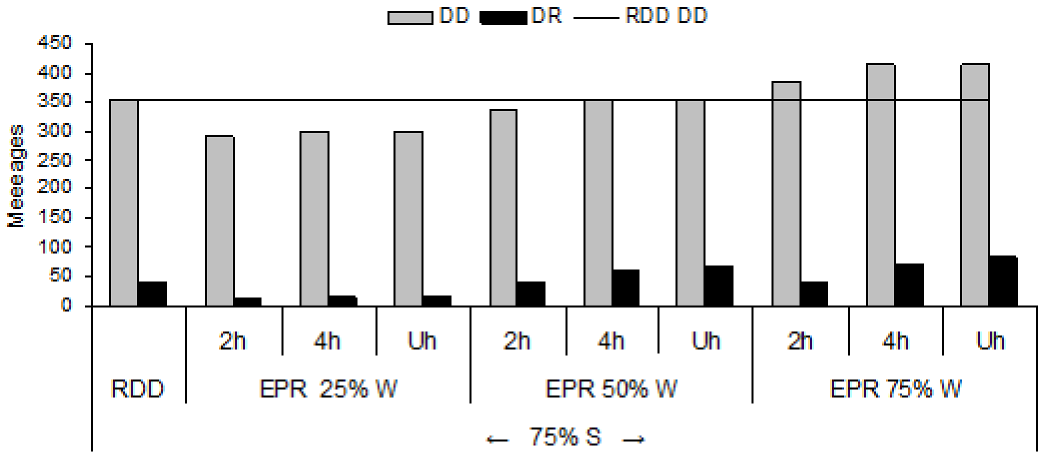


Figure 5.13: Performance in scenarios where MNs with $S(m) = C$ density is 75%.

5.11 Conclusion

In the first part of our evaluation process, the performance of the two repositories placement approaches (Betweenness and the Social Betweenness) was investigated in different scenarios where the density of mobile nodes and environment size are different. From the presented results we can conclude that SB performs better than B due to increased consistency of performance. The result clearly shows that placing the repository in the highest SB vertex always provide the highest data delivery rate therefore it can be considered is a good location for the repository.

In the second investigation, we have evaluated RDD performance across a range of metrics in comparison to the EPidemic Routing Protocol (EPR). Even with a single, well-placed repository, RDD is able to provide better performance when the density of mobile nodes that are willing to cooperate is 50% or lower. Another important finding is that RDD is able to significantly reduce DR, and consequently resource consumption. Our investigation also shows that when density of mobile nodes is low, RDD provide better performance in term of data delivery. As such results clearly indicate that RDD is a promising way to facilitate data dissemination in opportunistic networks especially in real life scenarios where other approaches fail due to the non cooperative nature of MNs in carrying data objects to other MNs.

Chapter 6

RDD Extensions

In this chapter, we introduce some extensions for RDD to investigate the effect of the initial assumptions. In the first section, we introduce Hybrid RDD (H-RDD) a modified version of RDD able to make use of any mobile node willing to relay. In the second section, we introduce Scalable RDD (S-RDD) a modified version of RDD targeting a bigger environment that needs more than one repository. Thus, a new strategy for repositories placement is introduced and the impact of the number of repositories used on the protocol performance is also investigated .0 Finally, we draw conclusions on H-RDD and S-RDD.

6.1 Hybrid RDD (H-RDD)

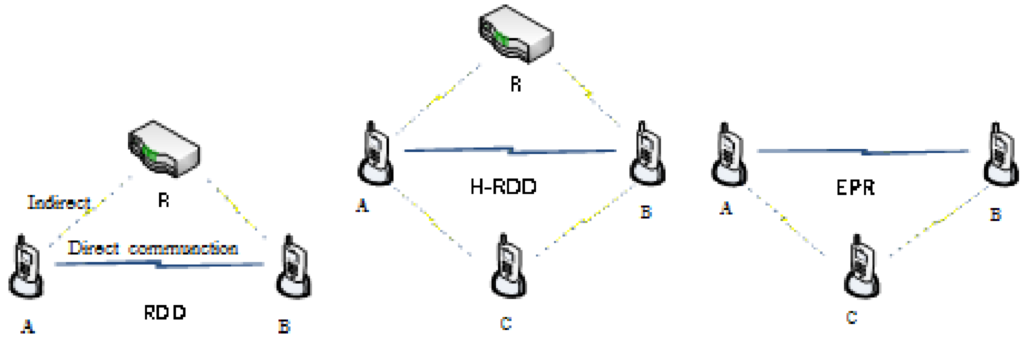
Although in chapter 4 we stated that mobile nodes are often not willing to relay, there is still the possibility of mobile nodes that are willing to relay. For example, this maybe encouraged through reputation systems or micro payment schemes.

As we highlighted in section 2.6, there are many reasons such as security concerns and limitations of mobile node resources, which may mean mobile users are not willing to provide relay services to other nodes. We believe an appropriate micropayment system such as that presented in [77], would allow a node to pay others who relay packets, motivating a mobile user to provide relay services.

Reputation is defined in [83] as the collected information about one entity's former behaviour as experienced by others. The reputation of the source plays an important role in information dissemination among mobile nodes. While finding the appropriate reputation schemes is beyond the scope of this research, however, employing the appropriate scheme could motivate mobile users to provide a relay service to those with high reputation value and not considered as security concerns for the encounter node.

In this section, we have introduced H-RDD which is able to take advantage of any relaying facilities that exist. We would expect this approach to prove a balanced alternative to RDD and EPR in terms of delivery and overhead. While RDD does not involve mobile nodes forwarding, H-RDD makes use of mobile nodes that are available and willing to relay in forwarding data objects. Figure 6.1 shows the possibility of communication between mobile nodes in H-RDD, RDD, and EPR. As we can see all support direct communication between mobile nodes while indirect communication employs repositories in RDD, mobile nodes in EPR, and both in H-RDD. Figure 6.1(a) shows that RDD allows mobile nodes (A and B) to communicate directly when encounter and indirectly by using the repository (R) node. Figure 6.1(b) shows that H-RDD allows mobile nodes (A and B) to communicate directly when encounter and indirectly trough another mobile node (c) and also through a repository node(R). Figure 6.1(c) shows that EPR allows mobile nodes (A and B) to communicate directly when encounter and indirectly trough another mobile node (c).

The design assumptions for H-RDD are the same as RDD apart from the second assumption “mobile nodes are not willing to relay” which we have removed in H-RDD. H-RDD is composed of the same RDD components which are: "data object management", "user preferences management", "data dissemination protocol", and "repository placement" as presented in chapter 4. While data object management, and repository placement are the same as in RDD, the other components are described below.



(a) RDD (b) H-RDD (c) EPR

Figure 6.1: Communication between MNs in RDD (a), H-RDD(b), and EPR(c)

6.1.1 User preference management

As before, each mobile node m subscribes to a set of channels $S(m)$ where $S(m) \subseteq C$, with the subscription process carried out manually by the device users. In addition to these channel subscriptions, in the design of H-RDD, we define a new user preference parameter called "willing to relay". If a mobile node m is willing to relay, the preference parameter $W(m)$ takes the value 1 and if m is not willing to relay takes the value 0 (Equation 6.1).

$$W(m) = \begin{cases} 1 & \text{if m willing to relay} \\ 0 & \text{if m not willing to relay} \end{cases} \tag{6.1}$$

The value for $W(m)$ is considered to be manually set by the users themselves, however this can be automated using the status of device available resources (such as the battery) to control switching this preference parameter on or off.

6.1.2 Data dissemination protocol

The data dissemination process is different in H-RDD because the mobile nodes can contribute in relaying data objects, which was not possible in RDD. The algorithm 6.1

shows the direct dissemination pseudocode while the algorithm 6.2 shows the indirect dissemination pseudocode.

6.1.3 H-RDD Evaluation

To evaluate H-RDD, simulation experiments were performed comparing H-RDD with both RDD and EPR. DD and DR were used as the performance metrics. The simulation setting used the same default values presented in Table 5.1 in Chapter 5 while the protocols (H-RDD, RDD and EPR) settings are presented in Table 6.1.

Parameter	*H-RDD*	*RDD*	*EPR*
Hop limit	Unlimited	n/a	Unlimited
Repository location strategy	SB	SB	n/a
$\|M\|$	25	25	25
$\|R\|$	1	1	0
$\|J\|$	25	25	25
$H(v)$	1	1	1
Density of MNs with $W(m) = 1$	50% (Scenario 1)	0%	50% (Scenario 1)
-	100% (Scenario 2)	0%	100% (Scenario 2)
-	100% (Scenario 3)	0%	100% (Scenario 3)
Density of MNs with $S(m) = C$	25% (Scenario 1)	25%	25% (Scenario 1)
-	50% (Scenario 2)	50%	50% (Scenario 2)
-	100% (Scenario 3)	100%	100% (Scenario 3)

Table 6.1: H-RDD Experiment setting

For the investigation, three simulation scenarios were considered (Table 6.1). The test results presented in Figure 6.2 and 6.3 are a comparison of performances between H-RDD, RDD and EPR for the same three simulation scenarios.

```
Initialization connection; t=Current time;
Send $S(m_i)$ and $W(m_i)$ from $m_i$ to $m_j$;
Send $S(m_j)$ and $W(m_j)$ from $m_j$ to $m_i$;
Let the list A be a copy of the elements of $|D_t^{wt}(m_i)|$ ordered by x(d);
Let the list B be a copy of the elements of $|D_t^{wt}(m_j)|$ ordered by x(d);
while Connection true do
    $A_{rec}$=false; $B_{rec}$=false;
    while Not $A_{rec}$ and $|A|$ >0 do
        d = Pop(A);
        if x(d) $<$ t then
            if c(d) $\in$ S($m_j$) then
                if $d \notin D_t^{rc}(m_j)$ then
                    Push $d$ from $m_i$ to $m_j$; $D_t^{rc}(m_j) = D_t^{rc}(m_j) \cup \{d\}$;
                    if W($m_j$) $= 1$ then
                        if $d \notin D_t^{W}(m_j)$ then
                            $D_t^{wt}(m_j) = D_t^{wt}(m_j) \cup \{d\}$;
                        end
                    end
                    $A_{rec}$=True;
                end
            else
                if W($m_j$) $= 1$ then
                    if $d \notin D_t^{wt}(m_j)$ then
                        $D_t^{wt}(m_j) = D_t^{wt}(m_j) \cup \{d\}$; $A_{rec}$=True;
                    end
                end
            end
        else
            $D_t^{wt}(m_i)$ = $D_t^{wt}(m_i) - \{d\}$; $A = A - \{d\}$
        end
    end
    while Not $B_{rec}$ and $|B|$ >0 do
        d = Pop(B);
        if x(d) $<$ t then
            if c(d) $\in$ S($m_i$) then
                if $d \notin D_t^{rc}(m_i)$) then
                    Push $d$ from $m_j$ to $m_i$; $D_t^{rc}(m_i) = D_t^{rc}(m_i) \cup \{d\}$;
                    if W($m_i$) $= 1$ then
                        if $d \notin D_t^{wt}(m_i)$ then
                            $D_t^{wt}(m_i) = D_t^{wt}(m_i) \cup \{d\}$;
                        end
                    end
                    $B_{rec}$=True;
                end
            else
                if W($m_i$) $= 1$ then
                    if $d \notin D_t^{wt}(m_i)$ then
                        $D_t^{wt}(m_i) = D_t^{wt}(m_i) \cup \{d\}$; $B_{rec}$=True;
                    end
                end
            end
        else
            $D_t^{wt}(m_j)$ = $D_t^{wt}(m_j) - \{d\}$; $B = B - \{d\}$
        end
    end
end
```

Algorithm 6.1: Direct dissemination pseudocode in H-RDD

```
Initialization connection;
t=Current time;
Send $S(m)$ and $W(m)$ from $m$ to $r$;
Let the list A be a copy of the elements of $|D_t^{wt}(m)|$ ordered by $x(d)$;
Let the list B be a copy of the elements of $|D_t^{wt}(r)|$ ordered by $x(d)$;
while Connection true do
    $A_{rec}$=false; $B_{rec}$=false;
    while Not $A_{rec}$ and $|A|$ >0 do
        d = Pop(A);
        if x(d) < t then
            if $d \notin D_t^{wt}(r))$ then
                Push $d$ from $m$ to $r$;
                $D_t^{wt}(r) = D_t^{wt}(r) \cup \{d\}$;
                $A_{rec}$=True;
            end
        else
            $D_t^{wt}(m)$ = $D_t^{wt}(m) - \{d\}$; $A = A - \{d\}$
        end
    end
    while Not $B_{rec}$ and $|B|$ >0 do
        d = Pop(B);
        if x(d) < t then
            if c(d) $\in$ S(m) then
                if $d \notin D_t^{rc}(m))$ then
                    Push $d$ from $r$ to $m$;
                    $D_t^{rc}(m) = D_t^{rc}(m) \cup \{d\}$;
                    if W(m) $= 1$ then
                        if $d \notin D_t^{wt}(m)$ then
                            $D_t^{wt}(m) = D_t^{wt}(m) \cup \{d\}$;
                        end
                    end
                    $B_{rec}$=True;
                end
            else
                if W(m) $= 1$ then
                    if $d \notin D_t^{wt}(m)$ then
                        $D_t^{wt}(m) = D_t^{wt}(m) \cup \{d\}$; $B_{rec}$=True;
                    end
                end
            end
        else
            $D_t^{wt}(r)$ = $D_t^{wt}(r) - \{d\}$; $B = B - \{d\}$
        end
    end
end
```

Algorithm 6.2: Indirect dissemination pseudocode in H-RDD

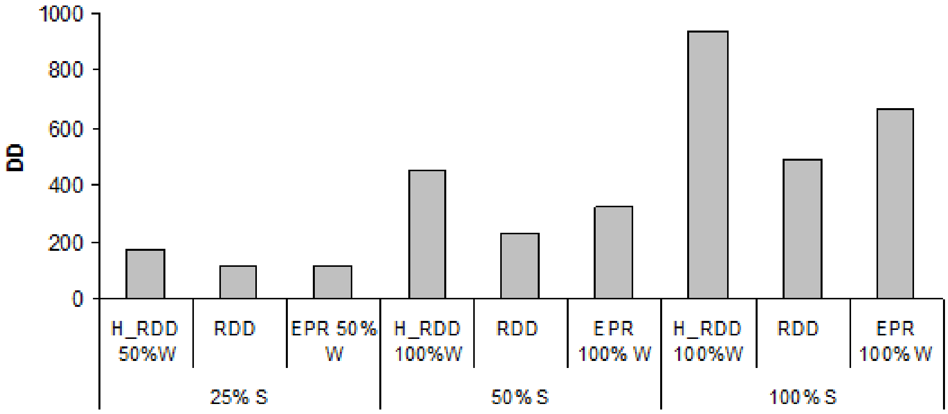


Figure 6.2: DD Performance in the three scenarios

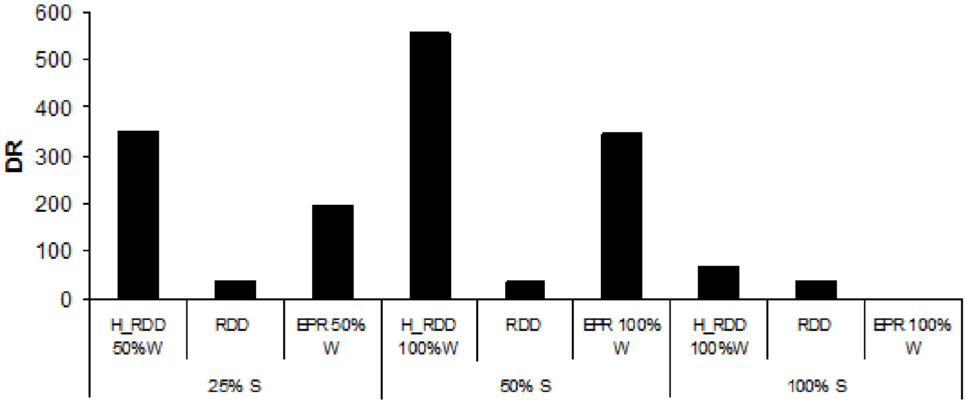


Figure 6.3: DR in the three scenarios

Figure 6.2 shows that H-RDD was able to improve DD approximately by 50% compared with RDD and EPR in the first scenario (25% of MNs subscribed and 50% of MNs willing to cooperate) but this improvement comes with a higher cost of message redundancy as we can see in Figure 6.3. The Figure also shows that DR in H-RDD always higher than DR in EPR. This occurs since indirect data dissemination routed through repository and mobile nodes require more than one hop,and some messages forwarded to repository or mobile nodes may never reach a final destination, while in

EPR the indirect data dissemination is routed only through mobile nodes.

In the second scenario (50% of MNs subscribed and 100% of MNs willing to cooperate), H-RDD was able to improve DD by approximately 90% compared to RDD and 40% compared to EPR. However, this improvement comes with a high cost of DR as we can see in Figure 6.3. In the third scenario (100% of MNs subscribed and 100% of MNs willing to cooperate), H-RDD was able to improve DD similar to the second scenario but with the lower DR as we can see in Figure 6.3.

6.2 Scalable RDD (S-RDD)

In Chapter 4 we assumed (assumption no 4) that the size of the target environments are campus or city region sized, where just one repository would be enough to cover the target area. In this section, we introduce S-RDD where this assumption has been removed, i.e. the size of target environment is bigger than the campus size and RDD needs more than one repository to cover the target area. To use S-RDD for any environment size, two questions need to be answered. First, where are the most suitable locations to deploy repositories? Secondly, how many repositories are needed to cover the target environment?

6.2.1 Repository location strategies

Since S-RDD targets an environment where more than one repository is needed finding the optimal locations for the repositories used is an essential part. Using SB locations ranking as the main strategy for finding the optimal locations for repositories could be a solution, but experimentation shows that using the SB locations ranking alone leads to choosing two locations that are very close to one another as seen in Figure 6.4. In this Figure, there is a graph composed of 100 vertices ($|J| = 100$) with the two repositories placed using the SB ranking.

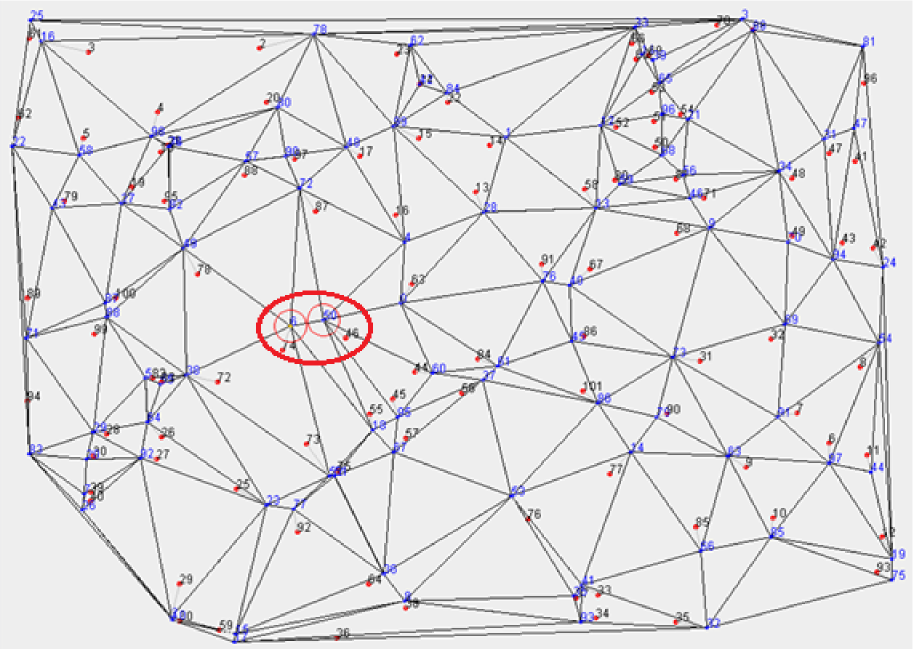


Figure 6.4: Two repositories located using SB location ranking

This finding goes with the fact that the highest ranking vertex or location are often surrounded by other vertices which also gain a high usage by the same people who are passing the highest ranking vertex, consequently these vertices are often get similar SB rankings. Because using this SB ranking strategy usually leads to placing repositories in close proximity, we propose Exclude SB (ExSB), based on measuring the SB of a vertex after excluding those vertices already chosen.

6.2.2 ExSB placement strategy

For each repository r, we define $l(r) \in J$ as the chosen location. L defines the set of locations where repositories are to be placed and is constructed greedily. Our placement strategy aims to find locations for more than one repositories. Please note that at start $L = \emptyset$. Equation 6.2 defines the $ExSB$ for vertex v.

$$ExSB(v, L) = \sum_{v_t \in V} \sum_{m \in M} \frac{\sigma_{h(m)v_t}(v, L).SpR(m, v_t)}{\sigma_{h(m)v_t}} \tag{6.2}$$

where $\sigma_{h(m)v_t}$ is the number of shortest paths from $h(m)$ to $v_t \in V$ and $\sigma_{h(m)v_t}(v, L)$ is the number of shortest paths from $h(m)$ to $v_t \in V$ that pass through a vertex v and do not pass any $l \in L$.

Since $L = \emptyset$ at the start, $ExSB$ is identical to SB for the first selection, i.e. r_1 is located at highest SB vertex. For subsequent repositories we calculate $ExSB$ for all vertices and we choose the highest $ExSB$ vertex to be the location for r_2 and so on. The pseudocode for $ExSB$ placement strategy is presented in algorithm 6.3

```
$N_R$ = number of repositories required ;
$L = \emptyset$;
while $|L| < N_R$ do
    Randomly choose $r \in arg_{v \in J-L}\ max\ ExSB(v, L)$;
    $L = L \cup l(r)$;
end
```

Algorithm 6.3: Placement repositories using ExSB

As can be seen in Figure 6.5, by applying the ExSB placement strategy for the same graph presented in Figure 6.5, a new location for the second repository have assigned and it's noticeable the new location is different from location in Figure 6.5 and the new location for the second repository is far from the first repository.

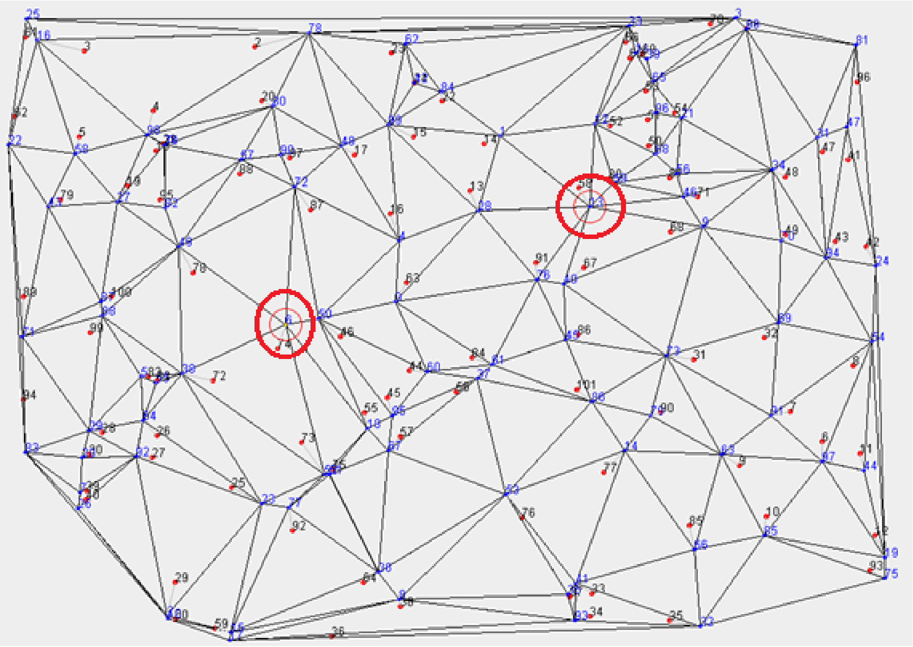


Figure 6.5: Two repositories located using ExSB location ranking

ExSB Evaluation To evaluate ExSB, simulation experiments were conducted comparing S-RDD performance using two different placement strategies which are SB and ExSB. The simulation setting used the same default values presented in Table 5.1 (in Chapter 5) while the protocol (S-RDD) settings are presented in Table 6.2.

Parameter	*S-RDD*
$\|M\|$	100
$\|R\|$	2
$\|J\|$	100
$H(v)$	1
R Placement strategy	Scenario 1 SB
-	Scenario 2 ExSB

Table 6.2: S-RDD Experiment setting

Figure 6.6 shows that using the ExSB strategy for placing repositories improves S-

RDD performance by approximately 5% compared with using SB. While the improvement is not huge, ExSB was able to place the repositories far from each other.

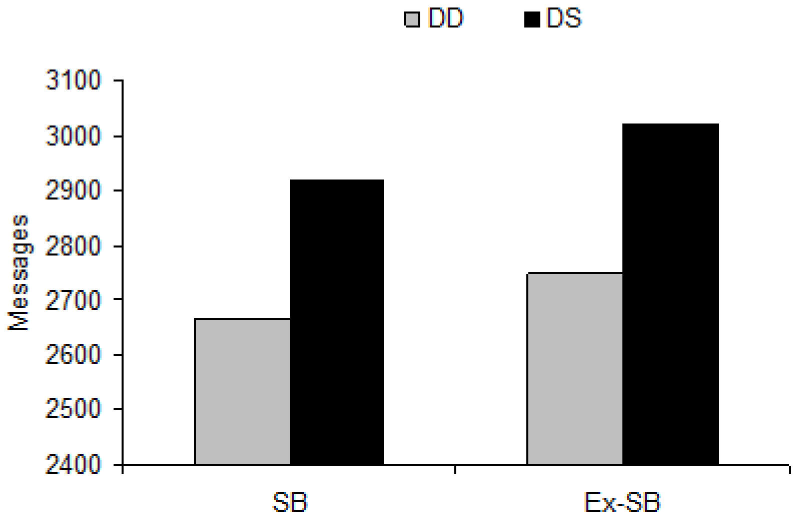


Figure 6.6: Comparison between SB and ExSB placement strategies

6.2.3 Number of Repositories needed in a given network

Adding more repositories to any given network should improve data delivery. However, a series of simulation experiments are carried out to investigate this postulate. The simulation setting used the same default values presented in Table 5.1 in Chapter 5 while the protocol settings are presented in Table 6.3.

Parameter	*Value*
$\|M\|$	25
$\|J\|$	40
$H(v)$	1
$\|R\|$	Scenario 1 $\|R\|$=0
-	Scenario 2 $\|R\|$=1
-	Scenario 3 $\|R\|$=2
-	Scenario 4 $\|R\|$=3
-	Scenario 5 $\|R\|$=4
-	Scenario 6 $\|R\|$=5
-	Scenario 7 $\|R\|$=6
-	Scenario 8 $\|R\|$=7

Table 6.3: Number of Repositories Experiment setting

The investigation showed that the effect of adding new repositories will decrease as more repositories are added, as seen in Figure 6.7. The big improvement in DD occurs when we add the 1^{st} repository and it is also noticeable that adding more repositories beyond 3 in the simulated scenarios has a minor effect on DD while adding more repositories beyond 6 have almost no effect on DD.

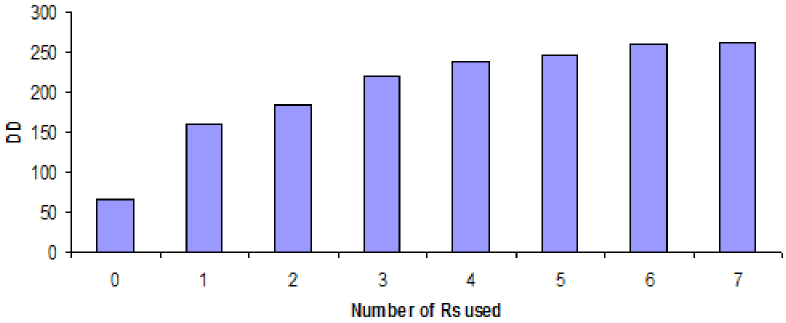


Figure 6.7: Impact of adding more repository on DD

6.3 Conclusion

In the first part of this chapter, we introduced and evaluated H-RDD, which extended RDD to be capable of using any forwarding opportunity from mobile nodes. Evaluation of H-RDD shows that it provides better DD even with 25% of mobile nodes cooperate. H-RDD can be seen as a balanced approach between approaches where relaying data objects depended only on MNs cooperation and the original RDD approach where relaying data objects depends only on repositories. One of the drawbacks of H-RDD lies in the high cost of data delivery.

In the second part, we have introduced and evaluated S-RDD. The main objective of S-RDD is extended RDD to be capable of dealing with an environment which needs more than one repository. One of the main problems associated with S-RDD is finding the appropriate placement strategy for more than one repository. To solve this problem, we have introduced ExSB, a placement strategy which employs the social betweenness in a modified form to ensure a geographic spread of repositories. As seen in the presented results in this chapter, ExSB was able to distribute the repositories in the simulated graph in more spread distribution and improve the DD.

In the final part of this chapter, we have investigated the effect of the number of repositories used on DD, calculating that the contribution of a new repository has an inverse relationship with number of repositories already placed in the network, i.e. the first repository has the highest impact of the network performance.

Chapter 7

Conclusions and future work

This thesis considers the problem of data dissemination in opportunistic networks in an environment where mobile devices are carried exclusively by humans. The main objective of this research was to design a data dissemination system capable of efficient data delivery to nodes using push-based communication. To achieve this goal, five research stages were conducted.

- **Background**. Thorough examination of the research background, research problem and related works (Chapter 2), concluded that the available solutions are mainly built around the assumption that mobile nodes are willing to relay data on behalf of other nodes, an assumption that is unlikely to be the case in practice. Understanding the nature and characteristics of mobile devices and their users was the driving force to reach our hypothesis (Chapter 1).
- **Evaluation**. After defining our hypothesis, an evaluation method was defined and marked the start of stage two (Chapter 3). Having studied all possible evaluation strategies, from building a real model to carrying out a simulation study, it was concluded that carrying out a simulation study was most appropriate for this research. To achieve this, a new Java-based network simulator (MdNS) was designed and built. Since the hypothesis was driven by characteristics of mobile devices and their users, and because the mobility pattern of mobile device users lies at the heart of these characteristics, therefore a mobility model which takes account of key characteristics of the user mobility was essential. A new "Human

mobility model" has been proposed, which is able to more closely mimic the real user's mobility, as was proven by conducting a series of experiments where the human mobility model was compared with real mobility traces.

- **Design**. Based on our hypothesis, we proposed Repository based data dissemination system (RDD) in Chapter 4. In RDD, two main questions needed to be answered: the first was how to determine the optimal location for the repository, and the second was how the system will be able provide push-based communication. Before attempting to answer those two questions, a set of assumptions was defined to simplify the design. The social betweenness strategy was introduced to evaluate the suitability of a location, which offered a better result than the betweenness strategy or a random choice. In RDD, users' data is classified into different channel "categories" upon data creation, and each mobile node should choose a set of channels to subscribe to. To answer the second question, our a new protocol employs this channel subscription to find out which data needs to be pushed to other nodes. Therefore channel subscription drives the pushing communication between nodes.
- **Simulation**. To evaluate RDD, extensive simulation experiments were conducted which mark stage 4 (Chapter 5). To improve the accuracy of the evaluation, 21 tests were carried out for each simulation experiment; this was achieved by using the Condor system provided by Cardiff University. The evaluation process included many aspects of RDD and the main finding of this research was that RDD is capable of providing an efficient and practical solution for data dissemination without the assumption that mobile nodes will always provide relay data on behalf of other nodes.
- **Extension**. In stage 5, some of the RDD extensions were introduced (Chapter 6). First extension was introduced Hybrid RDD (H-RDD), a modified version of RDD capable of use of any mobile nodes willing to relay. The conclusion we ob-

tained from conducting experiments was that H-RDD improve the performance comparing with RDD and EPR, which was expected because H-RDD employs mobile nodes in addition to the repository for relaying data objects. The second extension was introduced Scalable RDD, a modified version for RDD capable of use more than one repository. Within S-RDD we also introduced a new placement strategy for placing more than one repository. The results obtained from a series of experiment shows that S-RDD is able to provide good performance in a larger environment.

7.1 Thesis Contributions

This section summarizes our research contributions.

- We proposed a repository-based system for data dissemination in opportunistic networks, which we believe to be the first of its kind. Most of the existing protocols were exclusively built around the idea that mobile nodes are willing to relay on behalf of other mobile nodes, while our solution has been designed without this assumption to simulate the real life environment, where mobile nodes are usually not willing to relay. RDD enables mobile nodes to communicate with each other and share their data in an efficient manner without the needs for mobile node cooperation.
- In finding the optimal location for the repository, we have presented the social betweenness which adds a new dimension to the well-known betweenness centrality measure. While betweenness measures centrality only from a geographical point of view, social betweenness measures centrality from both the geographical and the human usage points of view. In addition to these two obvious dimensions, time and social relations are two other dimensions captured indirectly by social betweenness.

- For communication between devices, we proposed a push-based communication protocol that employs node channel subscription. Employing channel subscription enables the RDD dissemination protocol for deliver data objects to their interested nodes more efficiently.

- The evaluation process presented a new human mobility model designed to mimic exclusively human mobility patterns and its characteristics. All relevant human mobility pattern characteristics, such avoiding obstacles, human spatial preferences, choosing paths, and time modelling, were taken into account when our model was designed and implemented.

- In order to evaluate our approach, a dedicated network simulation tool was built. From designing and building our custom network simulator we have gained a lot of understanding and knowledge about network simulation tools. Our network simulation was built from three different integrated components: the simulation core, the mobility model, and the dissemination protocol. Building the simulation with modularity gave flexibility to our simulation.

- Extensive evaluation of the performance of RDD has been performed. In comparison with the epidemic protocol shows that RDD was able to achieve three goals. First, RDD was able to achieve a higher data delivery rate when the mobile nodes cooperation is less than 50%. Secondly, RDD significantly reduces data redundancy, and consequently resource consumption which is a crucial factor for mobile device's usability. Finally, RDD is able to provide a practical data dissemination solution because it easy to implement, does not need too much personal information, and cost effective.

- To extend RDD to able to make use of any available cooperation opportunity of mobile devices, a Hybrid RDD approach was proposed. Evaluation of H-RDD shows that H-RDD is able to provide better delivery rate compared to RDD and EPR. The most interesting feature of H-RDD as a data dissemination approach is that it does not rely completely on mobile nodes cooperation but still able to use any cooperation opportunity arise so it can be seen as a balanced practical solution.

7.2 Research outcome

The main research objective was to design an efficient data dissemination for opportunistic networks which reflects practical deployment issues. This research objective was successfully achieved by building the repository-based data dissemination system. The RDD building process started by employing a standalone fixed wireless device in the dissemination process, and was extended by employing the social dimension to find the optimal location for the repository, and finally the novel mobility model used in the evaluation process. The main advantage of the RDD is that it is practically well-suited to the mobile device characteristics and their users behaviour. Therefore, we strongly believe that many spatial information applications would benefit from using the RDD principles.

7.3 Directions for Future Research

To improve the performance of the system proposed in this thesis, the following future works are identified.

- In this thesis, we did not focus on security issues, as we have considered it to be outside the scope of this research. RDD was built with the attention of preserving user privacy, by avoiding sharing any data which describes an individual user beyond the next hop. A more advance techniques can be used, for example, using techniques similar to [52] it maybe possible to forward relevant messages, without explicitly sharing preferences. However, we believe providing privacy and security for the users is an essential part of network to stimulate user participation, therefore the user security remains open for future work.

- Scalable RDD uses many repositories, therefore mobile nodes communicate with all encountered repositories. To add more practicality of S-RDD, mobile nodes have to use a selection strategy to choose which repository to push their data objects into. Mobile nodes could employ channel history information to their decision for choosing a repository. Channel history for a device could count how many times a channel has been seen by a device. We believe using channel history would be an interesting direction to investigate how mobile nodes choose a repository. However deeper investigation of the problem is needed.

- Because the user mobility the duration of contact time between devices is usually short, therefore the number of data objects to push each direction limited. In this thesis, we use the expiration time of data object as the handling priority. The expiration time priority has been used in both directions, i.e. from mobile node to repository and from repository to mobile nodes therefore it would be interesting to investigate the use other priority method and use of a different method for each direction.

- In association of system deployment, many issues need to be clearly defined. Most important are defining the regulation of using repository services and defining a payment scheme for its use for any commercial purpose.

Bibliography

[1] Ian F Akyildiz, David M Gutierrez-Estevez, and Elias Chavarria Reyes. The evolution to 4G cellular systems: LTE-Advanced. *Physical Communication*, 3(4):217–244, 2010.

[2] Pentl Alex, , Fletcher Richard, and Hasson Amir. DakNet: Rethinking Connectivity in Developing Nations. *Computer*, 37(1):78–83, 2004.

[3] SM Allen, MJ Chorley, GB Colombo, E Jaho, M Karaliopoulos, I Stavrakakis, and RM Whitaker. Exploiting user interest similarity and social links for microblog forwarding in mobile opportunistic networks. *Pervasive and Mobile Computing*, 2011.

[4] L.S. Ashiho. Mobile Technology: Evolution From 1G To 4G. *Electronics for you*, 2003.

[5] Franz Aurenhammer. Voronoi diagrams a survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405, 1991.

[6] Maria Uden Avri Doria and Durga Prasad Pandey. Providing connectivity to the Saami nomadic community. In *2nd International Conference on Open Collaborative Design for Sustainable Innovation*, 2002.

[7] Rimon Barr and Zygmunt J. Haas. JiST/SWANS. Technical report, Wireless Networks Laboratory, Cornell University, 2005.

[8] C Boldrini, M Conti, and A. Passarella. ContentPlace: Social-aware data dissemination in opportunistic networks. In *11th international symposium on Modeling, analysis and simulation of wireless and mobile systems*, pages 203–210, 2008.

[9] Chiara Boldrini, Chiara Boldrini, Marco Conti, Jacopo Jacopini, and Andrea A. Passarella Andrea Passarella. HiBOp: A History based routing protocol for opportunistic networks. In *The IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WoWMoM 2007*, pages 1–12, 2007.

[10] Chiara Boldrini, Marco Conti, and Andrea Passarella. Design and performance evaluation of ContentPlace, a social-aware data dissemination system for opportunistic networks. *Computer Networks*, 54(4):589–604, 2010.

[11] Chiara Boldrini and Andrea Passarella. HCMM: Modelling spatial and temporal properties of human mobility driven by users social relationships. *Computer Communications*, 33(9):1056 – 1074, 2010.

[12] Franck Legendre Marcelo Dias de Amorim Borrel, Vincent and Serge Fdida. SIMPS: using sociology for personal mobility. *Networking, IEEE/ACM Transactions*, 17(3):831–842, 2009.

[13] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Ya Xu, and Haobo Yu. Advances in network simulation. *Computer*, 33(5):59 –67, 2000.

[14] John H Bryant. *Heinrich Hertz: The Beginning of Microwaves: Discovery of Electromagnetic Waves and Opening of the Electromagnetic Spectrum by Heinrich Hertz in the Years 1886-1892*. IEEE, 1988.

[15] B. Burns, O. Brock, and B. Levine. MV Routing and Capacity Building in Disruption Tolerant Networks. In *INFOCOM 2005, 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, volume 1, pages 398 – 408, 2005.

[16] A. Chaintreau, A. Hui, J. Crowcroft, C. Diot, R. Gass, and Scott J. Pocket switched networks: Real-world mobility and its consequences for opportunistic forwarding. *Technical Report UCAM-CL-TR-617, University of Cambridge, Computer Laboratory*, 2005.

[17] A. Chaintreau, A. Hui, J. Crowcroft, C. Diot, R. Gass, and Scott J. Impact of human mobility on the design of opportunistic forwarding algorithms. *IEEE Transactions on Mobile Computing*, 6:606–620, 2007.

[18] Xinjie Chang. Network simulations with OPNET. In *The 1999 Winter Simulation Conference*, volume 1, pages 307–314, 1999.

[19] L Paul Chew. Constrained delaunay triangulations. *Algorithmica*, 4(1-4):97–108, 1989.

[20] MJ Chorley, GB Colombo, MJ Williams, SM Allen, and RM Whitaker. Checking out checking in: observations on foursquare usage patterns. In *International Workshop on Finding Patterns of User Behaviors in NEtwork and MObility (NEMO)*, 2011.

[21] M. Conti, M. Mordacchini, and A. Passarella. Data dissemination in opportunistic networks using cognitive heuristics. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a*, pages 1 –6, 2011.

[22] S Corson and J Macker. RFC 2501: Mobile Ad Hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations, January 1999.

[23] P. Costa, C. Mascolo, M. Musolesi, and G.P. Picco. Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. *Selected Areas in Communications, IEEE Journal on*, 26(5):748 –760, 2008.

[24] Colleen Cuddy and Nancy R Glassman. Location-based services: Foursquare and gowalla, should libraries play? *Journal of Electronic Resources in Medical Libraries*, 7(4):336–343, 2010.

[25] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. An ANOVA test for functional data. *Computational statistics and data analysis*, 47(1):111–122, 2004.

[26] Alan Demers, Dan Greene, Carl Hauser, Wes Irish, John Larson, Scott Shenker, Howard Sturgis, Dan Swinehart, and Doug Terry. Epidemic algorithms for replicated database maintenance. In *Sixth annual ACM Symposium on Principles of distributed computing*, pages 1–12. ACM, 1987.

[27] Kevin Fall. A delay-tolerant network architecture for challenged internets. In *The 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 27–34, 2003.

[28] George S Fishman. *Principles of discrete event simulation*. John Wiley and Sons, New York, NY, 1978.

[29] George S Fishman. *Discrete-event simulation: modeling, programming, and analysis*. Springer, 2001.

[30] Reginald G. Golledge. Path selection and route preference in. human navigation: A progress report. *Working Paper, UCTC No. 277.*, (The University of California), 1995.

[31] Marta C. Gonzalez, Cesar A. Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 458(7235):238–238, 2009.

[32] Mark S Granovetter. The strength of weak ties. *American journal of sociology*, pages 1360–1380, 1973.

[33] R Gunther, L Levitin, B Schapiro, and P Wagner. Zipf's law and the effect of ranking on probability distributions. *International Journal of Theoretical Physics*, 35(2):395–417, 1996.

[34] Frank A Haight and Frank A Haight. *Handbook of the Poisson distribution*. Wiley New York, 1967.

[35] Timo Halonen, Javier Romero, and Juan Melero. *GSM, GPRS and EDGE performance: evolution towards 3G/UMTS*. Wiley, 2004.

[36] Klaus Herrmann. Modeling the sociological aspects of mobility in ad hoc networks. In *6th ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, MSWIM '03, pages 128–129, New York, NY, USA, 2003.

[37] Wei-jen Hsu, Kashyap Merchant, Haw-wei Shu, Chih-hsin Hsu, and Ahmed Helmy. Weighted waypoint mobility model and its impact on ad hoc networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(1):59–63, 2005.

[38] Wei-Jen Hsu, Thrasyvoulos Spyropoulos, Konstantinos Psounis, and Ahmed Helmy. Modeling spatial and temporal dependencies of user mobility in wireless mobile networks. *IEEE/ACM Transactions on Networking*, 17(5):1564–1577, 2009.

[39] Daniel Hughes, Geoff Coulson, and James Walkerdine. Free riding on Gnutella revisited: the bell tolls? *Distributed Systems Online, IEEE*, 6(6), 2005.

[40] Pan Hui, Augustin Chaintreau, Richard Gass, James Scott, Jon Crowcroft, and Christophe Diot. Pocket switched networking: Challenges, feasibility and implementation issues. *Autonomic Communication*, 3854:1–12, 2006.

[41] Pan Hui, Jon Crowcroft, and Eiko Yoneki. BUBBLE Rap: Social-based forwarding in delay-tolerant Networks. *IEEE Transactions on Mobile Computing*, 10:1576–1589, 2011.

[42] Verne Thompson Inman, Henry James Ralston, and Frank Todd. *Human walking*. Williams & Wilkins, 1981.

[43] Timur Friedman Vania Conan Jeremie Leguay. Evaluating mobyspace-based routing strategies in delay-tolerant networks. *Wireless communications and mobile computing*, 7(10):1171–1182, 2007.

[44] David B. Johnson and David A. Maltz. Dynamic source routing in ad hoc wireless networks. *Kluwer International Series in Engineering and Computer Science*, 353:153–181, 1996.

[45] Donald B Johnson. A note on Dijkstra's shortest path algorithm. *Journal of the ACM (JACM)*, 20(3):385–388, 1973.

[46] Gunnar Karlsson, Vincent Lenders, and Martin May. Delay-tolerant broadcasting. In *The 2006 SIGCOMM workshop on Challenged networks*, CHANTS '06, pages 197–204, New York, NY, USA, 2006.

[47] A. KERÄNEN. Opportunistic network environment simulator. Technical report, Helsinki University of Technology, Department of Communications and Networking, 2008.

[48] Jason LeBrun and Chen-Nee Chuah. Bluetooth content distribution stations on public transit. In *1st international workshop on Decentralized resource sharing in mobile computing and networking*, MobiShare '06, pages 63–65, New York, NY, USA, 2006.

[49] Jin-Shyan Lee, Yu-Wei Su, and Chung-Chou Shen. A comparative study of wireless protocols: Bluetooth, uwb, zigbee, and wi-fi. In *Industrial Electronics So-*

ciety, 2007. IECON 2007. 33rd Annual Conference of the IEEE, pages 46 –51, 2007.

[50] H. Lenando and R.M. Whitaker. The quality of spatial-temporal information from mobile peer-to-peer interactions. In *Wireless Telecommunications Symposium, 2008. WTS 2008*, pages 187–195, 2008.

[51] Gunnar Karlsson Lenders, Vincent and Martin May. Wireless ad hoc podcasting. In *Sensor, Mesh and Ad Hoc Communications and Networks (SECON'07)*, pages 273–283, 2007.

[52] Melek Onen Refik Molva M. J. Chorley Leontiadis, Iraklis and G. B. Colombo. Privacy preserving similarity detection for data analysis. In *Third International Conference on Cloud and Green Computing (CGC)*, pages 547–552, 2013.

[53] Min Li and Zhenjun Du. Dynamic social networking supporting location-based services. In *International Conference on Intelligent Human-Machine Systems and Cybernetics, 2009. IHMSC'09*, volume 1, pages 149–152, 2009.

[54] A. Lindgren, A. Doria, and O. Schelen. Probabilistic routing in intermittently connected networks. *Service Assurance with Partial and Intermittent Resources*, pages 239–254., 2004.

[55] Chen Ling Jyh. An evaluation of routing reliability in non-collaborative opportunistic networks. In *International Conference on Advanced Information Networking and Applications*, volume 0, pages 50–57, 2009.

[56] P. Lévy. *Théorie de l'addition des variables aleatories*. Gauthier-Villars, Paris, 1937.

[57] Matteo Marsili and Yi-Cheng Zhang. Interacting individuals leading to zipf's law. *Physical Review Letters*, 80(12):2741–2744, 1998.

[58] Mirco Musolesi and Cecilia Mascolo. A community based mobility model for ad hoc network research. In *2nd International Workshop on Multi-hop Ad Hoc Networks: From Theory to Reality*, pages 31–38, 2006.

[59] P. Nain, D. Towsley, Liu Benyuan, and Liu Zhen. Properties of random direction models. In *The 24th Annual IEEE International Conference on Computer Communications*, volume 3, pages 1897–1907, 2005.

[60] Jorge Nuevo. *A Comprehensible GloMoSim Tutorial.* INRS -Universite du Quebec, 2004.

[61] J. Osborne and A. Hardisty. Cardiff university condor pool: Background, case studies, and fec. In *UK e-Science All Hands Conference*, pages 361 – 365, 2006.

[62] Hui Pan, Chaintreau Augustin, Scott James, Gass Richard, Crowcroft Jon, Crowcroft Jon, and Diot Christophe. Pocket switched networks and human mobility in conference environments. In *ACM SIGCOMM workshop on Delay-tolerant networking*, pages 244–251, 2005.

[63] Hui Pan, A. Lindgren, and J. Crowcroft. Empirical evaluation of hybrid opportunistic networks. In *Communication Systems and Networks and Workshops*, pages 1–10, 2009.

[64] L. Pelusi, A. Passarella, and M. Conti. Opportunistic networking: Data forwarding in disconnected mobile ad hoc networks. *IEEE Communications Magazine*, 44(11):134–141, 2006.

[65] Charles E Perkins and Pravin Bhagwat. Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. *ACM SIGCOMM Computer Communication Review.*, 24(4):234–244, 1994.

[66] Juang Philo, Oki Hidekazu, Wang Yong, Martonosi Margaret, Peh Li Shiuan, and Rubenstein Daniel. Energy-efficient computing for wildlife tracking: design

tradeoffs and early experiences with ZebraNet. *ACM Sigplan Notices*, 37:96–107, 2002.

[67] Simon M. Reader, Matthew J. Bruce, and Susanne Rebers. Social learning of novel route preferences in adult humans. *Biology Letters*, 4:37–40, 2008.

[68] I. Rhee, M. Shin, S. Hong, K. Lee, and S. Chong. On the levy-walk nature of human mobility. In *Annual IEEE International Conference on Computer Communications*, pages 924–932, 2008.

[69] Injong Rhee, Minsu Shin, Seongik Hong, Kyunghan Lee, and Song Chong. Human mobility patterns and their impact on routing in human-driven mobile networks. In *The Sixth Workshop on Hot Topics in Networks (HotNets-VI)*, 2007.

[70] A. Ribeiro and R. Sofia. A survey of mobility models on wireless networks. Technical report, SITI, University Lusofona, 2011.

[71] Subir Kumar Sarkar, TG Basavaraju, and C Puttamadappa. *Ad hoc mobile wireless networks: principles, protocols and applications*. Auerbach Publications, 2007.

[72] James Scott, Jon Crowcroft, Pan Hui, Christophe Diot, et al. Haggle: A networking architecture designed around mobile users. In *WONS 2006: Third Annual Conference on Wireless On-demand Network Systems and Services*, pages 78–86, 2006.

[73] Chaoming Song, Zehui Qu, Nicholas Blumm, and .Albert-Laszlo Barabasi. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[74] Libo Song and David F. Kotz. Evaluating opportunistic routing protocols with large realistic contact traces. In *Second ACM workshop on Challenged networks*, pages 35–42, 2007.

[75] Le Tan and Liu Yong. On the capacity of hybrid wireless networks with opportunistic routing. *EURASIP Journal on Wireless Communications and Networking*, 2010.

[76] Small Tara and J. Haas Zygmunt. The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way). In *the 4th ACM international symposium on Mobile ad hoc networking & computing*, pages 233–244, 2003.

[77] Hitesh Tewari and Donal O'Mahony. Multiparty micropayments for ad hoc networks. In *Wireless Communications and Networking, 2003. WCNC 2003. 2003 IEEE*, volume 3, pages 2033–2040, 2003.

[78] Spyropoulos Thrasyvoulos, Spyropoulos Thrasyvoulos, Psounis Konstantinos, and S. Raghavendra Cauligi. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In *the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 252–259, 2005.

[79] Jing Tian, J. Hahner, C. Becker, I. Stepanov, and K. Rothermel. Graph-based mobility model for mobile ad hoc network simulation. In *Simulation Symposium, 2002. Proceedings. 35th Annual*, pages 337 – 344, 2002.

[80] Camp Tracy, Boleng Jeff, and Davies Vanessa. A survey of mobility models for ad hoc network research. *Wireless Communications and Mobile Computing*, 2(5):483–502, 2002.

[81] International Telecommunication Union. The World in 2013: ICT Facts and Figures. 2013.

[82] Amin Vahdat and David Becker. Epidemic routing for partially-connected ad hoc networks. Technical Report CS-200006, Duke University, 2000.

[83] Marco Voss, Andreas Heinemann, and Max Muhlhauser. A privacy preserving reputation system for mobile information dissemination networks. In *Security*

and Privacy for Emerging Areas in Communications Networks, 2005. SecureComm 2005. First International Conference on, pages 171–181, 2005.

[84] Yunsheng Wang and Jie Wu. Social-tie-based information dissemination in mobile opportunistic social networks. In *14th International Symposium on a World of Wireless, Mobile and Multimedia Networks*, 2013.

[85] Jorg Widmer and Jean-Yves Le Boudec. Network coding for efficient communication in extreme networks. In *The 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, pages 284–291, 2005.

[86] Matthew J Williams, Roger M Whitaker, and Stuart M Allen. Measuring individual regularity in human visiting patterns. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 117–122, 2012.

[87] Eiko Yoneki, Pan Hui, ShuYan Chan, and Jon Crowcroft. A socio-aware overlay for publish/subscribe communication in delay tolerant networks. In *10th ACM Symposium on Modeling, analysis, and simulation of wireless and mobile systems*, pages 225–234, 2007.

[88] X. L. Zhang, G. Neglia, J. Kurose, and D. Towsley. Performance modeling of epidemic routing. *Networking 2006: Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications Systems*, 3976:827–839, 2006.

[89] Wenrui Zhao, Mostafa Ammar, and Ellen Zegura. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *5th ACM international symposium on Mobile ad hoc networking and computing*, pages 187–198, 2004.

[90] Qian Zhang Zhensheng Zhang. Delay/disruption tolerant mobile ad hoc networks: latest developments. *the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, 7(10):1219–1232, 2007.

[91] Jinghui Zhong, Xiaomin Hu, Min Gu, and Jun Zhang. Comparison of performance between different selection strategies on simple genetic algorithms. In *International Conference on Computational Intelligence for Modelling, Control and Automation*, volume 2, pages 1115–1121, 2005.

[92] George Kingsley Zipf. *Selected studies of the principle of relative frequency in language*. Harvard University Press, 1932.

Appendix A

Zipf's law

Zipf's law usually refers to the size of an occurrence of an event relative to it's rank. The law is named after the linguist George Kingsley Zipf who first proposed it (1902-1950). In the 1930, George Kingsley Zipf observed that a few very frequent words make up a very large portion of any text or collection of texts, while the large majority of words occur relatively rarely [92].

Zipf proposed that if one arranged the words from a text in descending order of frequency, beginning with the most frequent word and continuing to the least frequent word in the text, one would find that the second word occurred roughly half as many times as the first word, and the third word occurred roughly one-third as often as the first word, and so on.

Zipf's law states that the size of the largest occurrence of an event z_i is inversely proportional to its rank and is defined by Equation (A.1).

$$z_i \propto \frac{1}{i^\alpha}(i = 1, ..., N) \tag{A.1}$$

where α is close to unity and N is the number of distinct occurrences of the event.

Zipf's law is most easily observed by plotting the data on a log-log graph, with the axes being log (rank order) and log (frequency).

Appendix B

Delaunay triangulation

To understand the Delaunay triangulation, which has been chosen to model the simulation environment, The Voronoi Diagram needs to be introduced. The Voronoi Diagram (VD) [5] is a ubiquitous structure (Figure B.1) that appears in a variety of disciplines such as biology, crystallography, etc. Given a set of sites (S) in the plane, determines the region of influence of each site this leads to the regions that decompose the plane, $VD(S)$ determines the region of influence of each site this leads to the regions that decompose the plane. Let S be a set of n sites in the plane. For each site $s \in S$, the voronoi region $VoR(s)$ is the set of all points that are at least as close to s as any other.

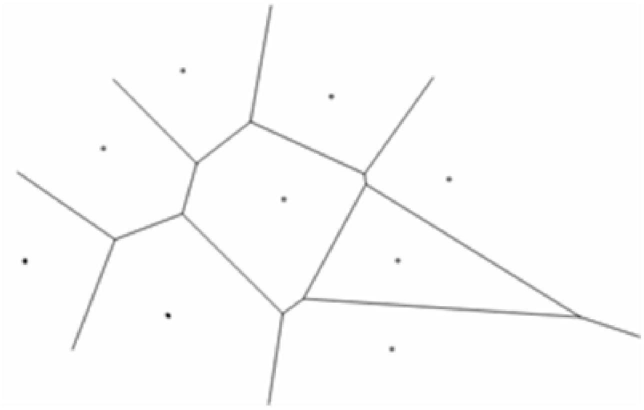


Figure B.1: Voronoi Diagram (VD)

The Delaunay triangulation $DT(S)$ [19], is the dual graph of the . This graph has a site for every voronoi region and has an edge e between two sites if the corresponding regions share an edge, and e_{sq} is an edge in $DT(S)$ if $VoR(s)$ and $VoR(q)$ share an edge. By connecting all edges, the shape will split into triangles as seen in Figure B.2.

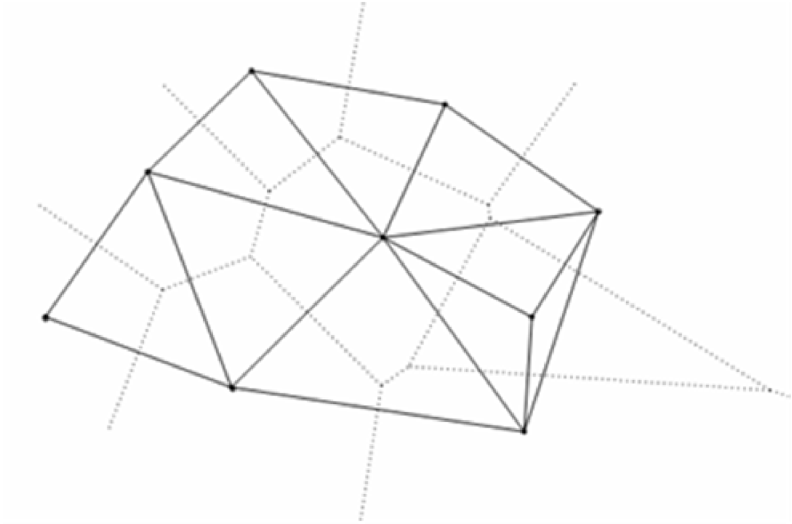


Figure B.2: Delaunay triangulation DT(S)