Universitat Autònoma de Barcelona

A methodological approach to routing protocol design and evaluation in opportunistic networking

Diego Mauricio Freire Bastidas

A methodological approach to routing protocol design and evaluation in opportunistic networking

Diego Mauricio Freire Bastidas

September, 2022

Version: My First Draft

Universitat Autònoma de Barcelona

Departament d'Enginyeria de la Informació i de les Comunicacions

Opportunistic Networks

A methodological approach to routing protocol design and evaluation in opportunistic networking

Diego Mauricio Freire Bastidas

Supervisors Dr. Carlos Borrego and Dr. Sergi Robles

September, 2022

Diego Mauricio Freire Bastidas

A methodological approach to routing protocol design and evaluation in opportunistic networking

Opportunistic Networks, September, 2022

Supervisors: Dr. Carlos Borrego and Dr. Sergi Robles

Universitat Autònoma de Barcelona

Departament d'Enginyeria de la Informació i de les Comunicacions

Plaça Cívica

08193 and Cerdanyola del Valles

Abstract

Over the years, wireless technologies have been integrated into everyday devices such as phones, tablets and laptops. At first, wireless communications were based on fixed infrastructure. Then, infrastructure-free networks were studied and implemented. These networks showed an opportunity for environments with dynamic topology, poor connectivity or disrupted communications.

Research showed that exploiting node-to-node communication opportunities enables information transmission in extreme network conditions, even when an end-to-end path may never exist. A typical network stack such as TCP/IP is not a feasible solution. Thus, some specific technologies have been developed to address these challenges. These emerging networking paradigms were Delay Tolerant Networks (DTN), a class of Opportunistic Networks (OppNets).

In this thesis, we have studied the OppNets environments and the characteristics that describe the intrinsic network behaviours. We have defined a set of characteristics that fully describes an OppNet scenario. The OppNet scenarios can be described with a vector of those characteristics.

Furthermore, we have designed a collection of scenarios, a corpus, for its use in the study and development of routing algorithms for Opportunistic Networks. To obtain these scenarios, we followed a methodology based on characterising the space and choosing the best exemplary items in such a way that the corpus as a whole was representative of all possible scenarios. Until now, research in this area used non-standard network traces, making it

challenging to evaluate algorithms and perform fair comparisons between them. These developments were hard to assess objectively and prone to unintentional biases that directly affected the quality of the research. Our contribution is more than a collection of scenarios; our corpus provides a fine collection of network behaviours that suit the development of routing algorithms, specifically evaluating and comparing them. If the scientific community embraces this corpus, the community will have a global-agreed methodology where the validity of results would not be limited to specific scenarios or network conditions. Thus, avoiding self-produced evaluation setups, saving time, availability problems and selection bias. New research in the area will be able to validate the routing algorithms already published. It will also be possible to identify the scenarios that suit better for specific purposes, and results will be easily verified. The corpus is available free to download and use.

Resumen

A lo largo de los años, las tecnologías inalámbricas se han integrado en dispositivos cotidianos como teléfonos, tabletas y ordenadores portátiles. En un principio, las comunicaciones inalámbricas se basaban en una infraestructura fija. Después se estudiaron e implementaron redes libres de infraestructura. Estas redes mostraron una oportunidad para entornos con topología dinámica, mala conectividad o comunicaciones interrumpidas.

La investigación demostró que aprovechar las oportunidades de comunicación de nodo a nodo permite la transmisión de información en condiciones de red extremas, incluso cuando es posible que nunca exista una ruta de extremo a extremo. Una pila de red típica como TCP/IP no es adecuada. Por lo tanto, se han desarrollado algunas tecnologías específicas para hacer frente a estos desafíos. Estos paradigmas de redes emergentes fueron las redes tolerantes al retraso (DTN), un tipo de Redes Oportunistas (OppNets).

Hemos estudiado los entornos OppNets y las características que describen los comportamientos intrínsecos de la red. Hemos definido un conjunto de características que describen completamente un escenario de OppNet. Los escenarios de OppNet se pueden describir con un vector de esas características.

Además, hemos diseñado una colección de escenarios, un corpus, para su uso en el estudio y desarrollo de algoritmos de enrutamiento para Redes Oportunistas. Para la obtención de estos escenarios se siguió una metodología basada en la caracterización del espacio y la elección de los mejores ejemplares de manera que el corpus en su conjunto fuera representativo de todos los escenarios posibles. Hasta ahora, la investigación en esta área utilizaba trazas de red no estándar, lo que dificultaba la evaluación de algoritmos y la realización de comparaciones justas entre ellos. Estos desarrollos fueron difíciles de evaluar objetivamente y propensos a sesgos no intencionales que afectaron directamente la calidad de la investigación. Nuestra contribución es más que una colección de escenarios; nuestro corpus proporciona una

excelente colección de comportamientos de red que se adaptan al desarrollo de algoritmos de enrutamiento, evaluándolos y comparándolos específicamente. Si la comunidad científica adopta este corpus, la comunidad tendrá una metodología acordada globalmente donde la validez de los resultados no se limitará a escenarios específicos o condiciones de red. Así, se evitan montajes de evaluación autoproducidos, ahorro de tiempo, problemas de disponibilidad y sesgos de selección. Nuevas investigaciones en el área podrán validar los algoritmos de enrutamiento ya publicados. También será posible identificar los escenarios que mejor se adapten a propósitos específicos, y los resultados se verificarán fácilmente. El corpus está disponible para su descarga y uso de forma gratuita.

Resum

Al llarg dels anys, les tecnologies sense fils s'han integrat en dispositius quotidians com ara telèfons, tauletes i portàtils. Al principi, les comunicacions sense fil es basaven en una infraestructura fixa. Després, es van estudiar i implementar xarxes sense infraestructura. Aquestes xarxes van mostrar una oportunitat per a ambients amb topologia dinàmica, una mala connectivitat o comunicacions interrompudes.

La investigació va mostrar que l'explotació de les oportunitats de comunicació node-a-node permet la transmissió d'informació en condicions extremes de xarxa, fins i tot quan un camí d'extrem a extrem no pot existir. Una pila de xarxa típica com TCP/IP podria no ser factible. S'han desenvolupat algunes tecnologies específiques per abordar aquests reptes. Aquests paradigmes emergents de xarxa van ser Delay Tolerant Networks (DTN), una classe de Opportunistic Networks (OppNets).

Hem estudiat els entorns OppNets i les característiques que descriuen els comportaments de les xarxes intrínseques. Hem definit un conjunt de característiques que descriuen plenament un escenari OppNet. Per tant, els escenaris d'OppNet es poden descriure amb un vector d'aquestes característiques.

A més, hem dissenyat una col·lecció d'escenaris, un corpus, per al seu ús en l'estudi i el desenvolupament d'algorismes d'encaminament per a Xarxes Opportunistes. Per a obtenir aquests escenaris, seguim una metodologia basada en la caracterització de l'espai i l'elecció dels millors elements exemplars de manera que el corpus en el seu conjunt sigui representatiu de tots els escenaris possibles. Fins ara, la recerca en aquest àmbit utilitza traces de xarxa no estàndard, fent que sigui difícil avaluar algoritmes i realitzar comparacions justes entre ells. Aquests desenvolupaments van ser difícils d'avaluar objectivament i propensos a biaixos no intencionats que van afectar

directament la qualitat de la recerca. La nostra contribució és més que una col·lecció d'escenaris; el nostre corpus proporciona una col·lecció fina de comportaments de xarxa que s'adapten al desenvolupament d'algorismes d'encaminament, en particular l'avaluació i comparació d'aquests. Si la comunitat científica accepta aquest corpus, la comunitat tindrà una metodologia d'acord global on la validesa dels resultats no es limitarà a escenaris específics o condicions de xarxa. Així, evitant les configuracions d'avaluació autoproduides, estalviant temps, problemes de disponibilitat i biaix de selecció. Una nova investigació a la zona podrà validar els algoritmes d'encaminament ja publicats. També serà possible identificar els escenaris que s'adaptin millor per a fins específics, i els resultats es verificaran fàcilment. El corpus està disponible gratuïtament per descarregar i utilitzar.

Acknowledgement

I have always considered that multiple actors are involved in education, some more visible than others. For that reason, I want to thank the people that have made it possible to achieve this goal. It is not my intention to turn this into a list because there are many, and naming all would not be possible. However, I think it is fair to put some names that accompanied me on this path.

First, I want to thank my dear wife, Paulina, for believing in me and for the effort and tenacity that have led me to complete this doctorate. I admire you.

I want to thank my parents, Raul and Sylvia, for all their unconditional love, support and help towards my wife and me. Thanks to my family, which is large and naming them would take me another thesis. Thanks to all of you.

The warmest thanks to my supervisors, Dr Carlos Borrego and Dr Sergi Robles, for their help, which has gone beyond academics. All these years, I have felt sincere support and consideration from you. A fraternal thanks for your caring, advice and guidance. Thanks to Carlos for speaking to me frankly and directly, for being aware that this thesis progresses at the pace it should and for helping me with the corrections. And to Sergi, a "thank you" is not enough to express my gratitude and appreciation. In times of despair, I always found his help, guidance and words of encouragement. Thank you also for allowing me to teach in the dEIC department. It has been an enriching experience.

At the same time, I would like to thank the Department of Engineering of Information and Communications (dEIC), a department with a pleasant environment to work and talk about academic and non-academic matters. The conversations after lunch were always enriching and uncomplicated, especially with Adrià, Ramon, Cristina, MCarmen and Ian.

I would like to thank my cousins Sandra and Ramon. Their calls have accompanied me on this journey.

And my dear friends Aïda and Arnau for that space of laughter and sincere friendship. Our meetings always were fun, and the games and the conversations filled my heart. To my friends Francisco and Daniela, a heartfelt thank you for your help over the years. To my friend Iván who has always been there to talk.

Thank you all. I have been lucky to have met such good people.

List of Figures

List of Tables

List of Algorithms

Contents

Part I

Preliminaries

1 Introduction

> *We need to understand that if we all work on inclusion together, it's going to be faster, broader, better, and more thorough than anything we can do on our own.*
>
> — **Ellen Pao**
> (former Reddit CEO)

1.1 Introduction

Over the years, wireless communications have provided connectivity solutions when other technologies struggle to do so. They have proven to be valuable assets for transmitting information and have constantly been evolving at a fast pace. Wireless communications provide a new connectivity approach with implicit mobility.

In the last decade, wireless technologies have been integrated into everyday devices such as phones, tablets and laptops. At first, wireless communications were based on fixed infrastructure. Then, researchers started to investigate whether or not it is possible to transmit information without fixed infrastructure. Such infrastructure-free networks were possible and showed an opportunity for environments with dynamic topology, poor connectivity or disrupted communications.

Research showed that exploiting node-to-node communication opportunities enables information transmission in such extreme network conditions even when end-to-end path may never exist. Using a typical network stack such as TCP/IP might not be feasible. Thus, some specific technologies have been developed to address these challenges. These emerging networking paradigms were known as Delay Tolerant Networks (DTN) or Opportunistic Network (OppNet). For the rest of this thesis, the term OppNet(s) refers to an Opportunistic Network(s).

Research in routing algorithms might be the most exciting topic in OppNets. Routing algorithms can be understood as the intelligence behind the decisions that enable communication in such challenged networks. In OppNets, routing algorithms aim to deliver messages from source to destination nodes without using fixed infrastructure. Over the years, many authors have proposed routing algorithms advocating for efficient communication. However, the evaluation and comparison methods have not been agreed upon across the scientific community, making it difficult to replicate and contrast routing contributions.

The applications of OppNets have been widely studied, and unlike traditional networks, the OppNet scenarios could be very different from one to another, and have been poorly described in the literature until now. This thesis will study and characterise the features that describe an OppNet scenario.

This research work is focused on the study of how the evaluation and comparison of routing algorithms for OppNets. In this thesis we analyse the main problems that hinders the evaluation methodology and thus, affecting the objective quality. We are confident that this thesis' contribution will enhance the evaluation and comparison of routing algorithms.

1.2 Objectives

This thesis aims to enhance the consistency of the evaluation and comparison routing algorithms for Opportunistic Networks. The cornerstone of this study is contributing to the standardisation of the evaluation methodology. To achieve the main objective, the following specific objectives have been proposed:

- To analyse the elements involved that constitute an OppNet
- To evaluate the evaluation and comparison methodology
- To determine the characteristics that shape the OppNet scenario
- To propose a new methodology for evaluating and comparing OppNet routing algorithms, considering coverage, scope, quality and usability.
- To design and implement a general purpose architecture to evaluate and compare OppNet routing algorithms fairly.

1.3 List of contributions

The contributions produced by this thesis are listed here:

- A design of a creation methodology of routing algorithms.
- A design and implementation of an evaluation methodology of routing algorithms.
- The definition and characterisation of an OppNet scenario.
- An implementation for the creation of synthetic OppNet scenarios.
- A design and implementation of a methodology that creates and select a collection of representative OppNet scenarios for the evaluation and comparison of routing algorithms.

1.4 List of publications

The publications produced by this thesis are listed here:

- Freire, Diego; Robles, Sergi; Borrego, Carlos. "Towards a Methodology for the Development of Routing Algorithms in Opportunistic Networks". In: *The Sixteenth International Conference on Wireless and Mobile Communications ICWMC 2020*, Oporto, Portugal. 19 October 2020.

- Freire, Diego; Borrego, Carlos; Robles, Sergi. "Corpus for Development of Routing Algorithms in Opportunistic Networks". In: *Applied Sciences 12.18* (2022), p. 9240.

1.5 Thesis Structure

Chapter 2

This chapter describes the related work that will help the reader to understand the thesis. As previously stated, the main objective of this thesis is to enhance the evaluation and comparison of routing algorithms in Opportunistic Networks. The Opportunistic Networking paradigm, the routing protocols and the state-of-the-art evaluation process are explained. It review the concepts used in literature to describe OppNet scenarios. This chapter will also present some insights from other research fields when evaluating and comparing algorithms. More precisely, three different research fields will help clarify the ideas supporting this work. Those research fields are data compression, speech recognition, and linguistics.

Chapter 3

This chapter introduces a methodology for the development of routing algorithms that take into consideration opportunistic networking. The proposal focuses on the methodology's rationale and highlights its most important stages and components. It also discusses the importance of two core elements in the process of protocol designing: scenario selection based on essential characteristics and the choice of standard evaluation metrics.

Chapter 4

This chapter defines a complete methodology for developing a corpus and presents a corpus for evaluating and comparing routing algorithms. This chapter presents the corpus creation methodology with five well-delimited stages, each with specific inputs, outputs and tasks. The result of this chapter is the main proposal of this thesis, a collection of scenarios named corpus.

Chapter 5

Chapter five present the conclusions that this thesis has produced and describes the future research lines.

2 Related Work

> *A picture is worth a thousand words. An interface is worth a thousand pictures.*
>
> — **Ben Shneiderman**
> (Professor of Computer Science)

This chapter describes the related work that will help the reader to understand the thesis. As previously stated, the main objective of this thesis is to enhance the evaluation and comparison of routing algorithms in Opportunistic Networks. The Opportunistic Networking paradigm, the routing protocols and the state-of-the-art evaluation process are explained. It review the concepts used in literature to describe OppNet scenarios. This chapter will also present some insights from other research fields when evaluating and comparing algorithms. More precisely, three different research fields will help clarify the ideas supporting this work. Those research fields are data compression, speech recognition, and linguistics.

2.1 Opportunistic networks (OppNets)

Opportunistic network(s) (OppNet(s)) are wirelessly connected devices that exchange information by exploiting connection opportunities. Devices from these networks have embedded wireless capabilities, such as smartphones, tablets, and smartwatches. Devices seize opportunities to connect with other devices within reach of their wireless capabilities in an opportunistic manner [1]. OppNets allow information exchange among devices even when an end-to-end path may never exist [2]. Moreover, the network's topology variations are considered normal in OppNets, and it is due to the wireless nature of the devices [3]. Finally, devices are commonly known as nodes.

Contrary to traditional networks, OppNets do not rely on fixed infrastructure to exchange information among nodes [4]. The nonexistence of a fixed infrastructure raises problems such as long or variable delays, connectivity disruptions and asymmetric data rates, which are considered normal due to the implicit mobility that nodes have [5].

OppNets use the store-carry-and-forward paradigm to transmit information in such a challenging environment. A node implementing the store-carry-and-forward paradigms stores messages until another node offers a suitable communication opportunity to transfer the message [6, 7]. This node-to-node transference of messages will eventually reach the intended destination node.

In addition, OppNet applications are suitable to provide effective communication solutions in challenged environments such as cellular network offloading [8], communication in challenged areas [8], censorship circumvention [8], mobile *ad hoc* social networks [2], offline social networks [2], Internet of Vehicles [9], Information-Centric Networking [10], underwater communications [11], proximity-based applications [2], among others.

2.2 Routing algorithms in OppNets

Routing algorithms are the intelligence allowing information transmission between the source and destination node(s). It can also be described as implementing a routing function that integrates the store-carry-and-forward paradigm. A routing algorithm aims to maximise message delivery and minimise network resource consumption. Unfortunately, unlike traditional networks such as the Internet, OppNet routing algorithms strategies deliver more than a single copy of the message when nodes are trying to reach the destination. Thus, managing to operate in this network where disconnection and delay are considered normal behaviour.

Researchers have been keen on the development of new routing algorithms for OppNets. One of the principal differences among routing algorithms depends on the strategy implemented. Works like [12] and [13] provide a complete list of the routing algorithms for OppNets. The reader can find in work [13] how routing algorithms can be classified depending on the implemented strategies.

Some routing forwarding strategies implement different dissemination strategies such as an epidemic [14], probabilistic [15], number of copies [16], or based on neighbourhood contact history [17] to deliver messages among nodes.

Besides the forwarding decisions, the routing algorithm must decide the best candidate(s) to receive a message among all available nodes and update the message lifetime. Another directive that routing algorithms are in charge of is whether a message should be stored or deleted.

2.3 Characteristics and metrics in OppNets

This section explains the use of characteristics and metrics in OppNets and why they are fundamental concepts for this thesis. This section also describes why characteristics and metrics are closely related and how the characteristics influence the metrics.

2.3.1 Characteristics

Characteristics are descriptors that help to characterise OppNets. As mentioned in Section 2.1, OppNets environments are heterogeneous, and with such heterogeneity among OppNets, it is tempting to say a priori that one OppNet is different to another. However, that would not be the general case. As said in [18], characteristics are deployment facts expressed in numbers for a network. Therefore, one way to establish the difference is by characterising the OppNets, which means numerically expressing the singular features of a particular OppNet. The result of the characterisation is a set of meaningful characteristics that will describe the uniqueness of an OppNet. Until now, the literature reviewed showed no characterisation of OppNets.

2.3.2 Metrics

Metrics are quantitative information that a measurement function retrieves, and they are usually numerical values that can be interpreted as the degree of occurrence of an attribute [19]. Metrics are extensively used among researchers seeking to evaluate, compare or measure some behaviours that an OppNet routing algorithm has [20, 21]. For example, metrics can be used to assess a comparison between all messages created versus the number of messages delivered to their intended destinations.

There are some metrics that most OppNet researchers tend to use to prove performance hypotheses. Among them, three stand out because of their presence in most works related to OppNets: delivery ratio, delivery delay, and delivery cost [12]. However, some authors do not use these metrics but modify versions of them to fit certain hypotheses. In other cases, some authors even find it necessary to establish entirely new metrics to test their work [22, 23]. Table 2.1 shows the frequently used metrics. Metrics have been classified by their computing data source into four major classes: energy consumption, media occupancy, network and messages. Metrics shown in Table 2.1 is not an exhaustive list.

2.3.3 Relationship

Metrics and characteristics are closely related. The relationship is given because of the routing algorithm working within an OppNet. On the one hand, characteristics describe a particular OppNet where a particular routing

Table 2.1.: Metrics for performance evaluation on OppNets routing algorithms classified by their computing source, with four major classes: energy, media, network and messages.

Consumption	Energy consumption Wireless card consumption Node energy level Sensor lifetime Node lifetime Core usage Energy dissipated
Media	Channel utilization Partial transmissions reduction Data carried in media Throughput (kbits/s) Throughput overhead

Message	Delivered Relayed Aborted Droped Delivery probability Overhead ratio Latency Hop-count Buffertime
Network	Delivery ratio End to end delay Medium delay Delay overhead Control packets Control packets overhead

algorithm will transmit information. On the other hand, the particular routing algorithm will behave in a specific fashion. Behaviour that the metrics will measure. Therefore, changes in the OppNet characteristics will affect the metrics obtained. It is a one-way relationship since metrics cannot influence nor modify the characteristics of the OppNet.

2.4 Evaluation and comparison of OppNet routing algorithms

This section describes the evaluation and comparison of routing algorithms, and why they have been performed in the literature.

The efficiency and performance of a network depend entirely on the routing algorithm. OppNet configurations can be complex and diverse. For example, node distribution can be either sparse or dense and still be an OppNet. Since the forwarding of messages relies on the routing algorithm and thus OppNet performance, an accurate measurement of the performance is a must.

Evaluating and comparing a routing algorithm can be described as an inter-technique and intra-technique, respectively [24]. The first, the evaluation,

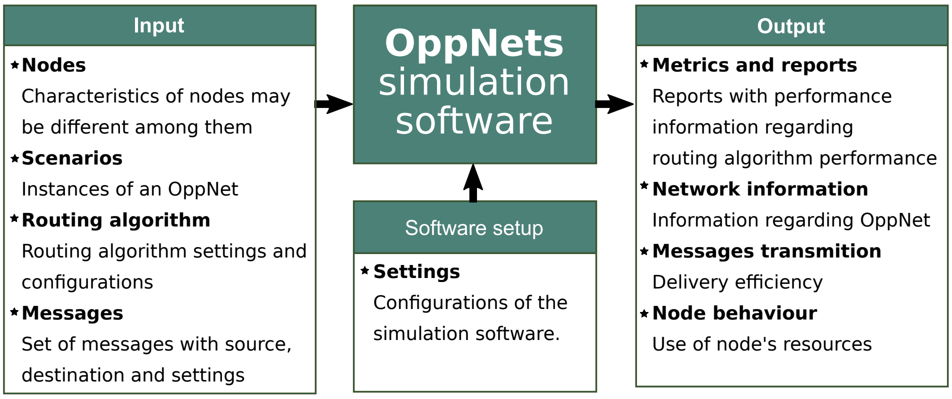


Figure 2.1.: Scenario's input, output and configuration elements that enable an OppNet simulation

quantitatively recognises the routing algorithm behaviour for a particular OppNet environment. The evaluation process of one does not require other(s) routing algorithms to assess their performance. The second one, the comparison, ranks the performance of a routing algorithm against other(s) routing algorithms.

2.4.1 Evaluation

Evaluation can be defined as a quantification of a service by a communication system, and it uses metrics with information regarding the behaviour to evaluate. In a routing algorithm for OppNets, different behaviours can be evaluated. For example, message delivery ratio or energy efficiency can be evaluated. Evaluating one or the other will depend on metrics containing information regarding the delivery of messages or energy consumption, respectively.

Evaluating the performance of a routing algorithm requires clear definitions of the what and how. The "what" are the objectives, while the "how" are the resources needed to obtain an evaluation. For example, suppose the objective is to evaluate the number of messages delivered by the nodes. In that case, that objective is the "what", while the metrics, devices, and configuration acting together to retrieve the information are the "how" [25].

There are several techniques to evaluate the performance of routing algorithms. The first way is by measuring the behaviour from real-world or testbed implementations. This type of deployment can be costly and sometimes difficult. The second way is using simulation tools which mimic the elements and their interactions in the real world. The third way is using analytical models. This approach creates a mathematical model of the system to evaluate. An analytical approach usually requires simplifications and assumptions about the system [25].

However, the second evaluation alternative of performance evaluation, the simulation, is prominent in the reviewed literature. Implementing a simulation could be considered easier than real testbeds but also offers control over the elements involved in a simulation. The elements to deploy an OppNet simulation system have three main components, which are depicted in Figure 2.1, discussed in more detail in Section 2.5. The first element, the input, is all the information to replicate an OppNet that transmits messages among the nodes. The second element, the software configuration, is the configuration to run the simulation software. Moreover, the third and final element, the output, is the information that the simulation retrieves after the simulation(s) is(are) performed.

As the reader can appreciate from Figure 2.1, scenarios and metrics are involved in the simulation of OppNets. On the one hand, there are several traces suitable for OppNets. Sites like CRAWDAD [26] gather mobility traces datasets that are shared among the scientific community. Some studies tend to frequently use some traces known in the community as "well-known traces" [27] or "well-known scenarios" [28, 29]. Datasets such as Asturias [30], Taxis Roma [31], Taxis San Francisco [32] and Cambridge/Haggle [33] are some of those that are usually included in literature as "well-known traces".

Similarly, some metrics are used more frequently than others in order to evaluate the performance of a routing algorithm. Among them are the delivery ratio, delivery delay, and network overhead [13, 25].

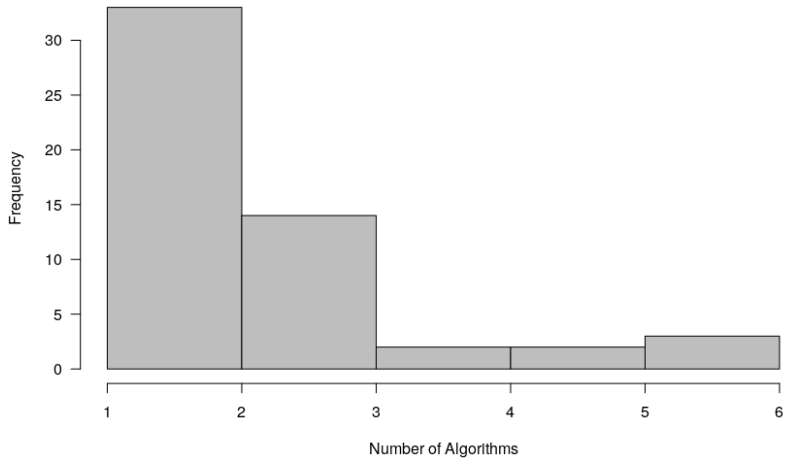


Figure 2.2.: Number of routing algorithms used in scientific articles for comparison purposes

2.4.2 Routing performance comparison

Providing a comparison is an important element in the design of new routing algorithms. Research without a comparison might lower the chances of being accepted to be published [34]. That might be why comparing OppNet routing algorithms with peers is commonly presented in scientific articles. Most of the time, authors tend to use a set of selected algorithms that will be used to establish a rank among them. There are some practices that it is important to highlight as follows.

Figure 2.2 depicts the number of routing algorithms used when comparing them. The horizontal axe is the number of routing algorithms ascended-ordered within [1-6], while the vertical axe is the frequency of occurrence. In Figure 2.2, it is shown that approximately 62% of the routing algorithm contributions use up to two routing algorithms when comparing their proposals. Among the routing algorithms frequently used when comparing are Epidemic, Spray & Wait, MaxProp, and BubleRap [34].

Figure 2.3 shows the routing algorithms and their comparisons using a cloud graph. In Figure 2.3, a cloud graph is constituted by vertices and edges.

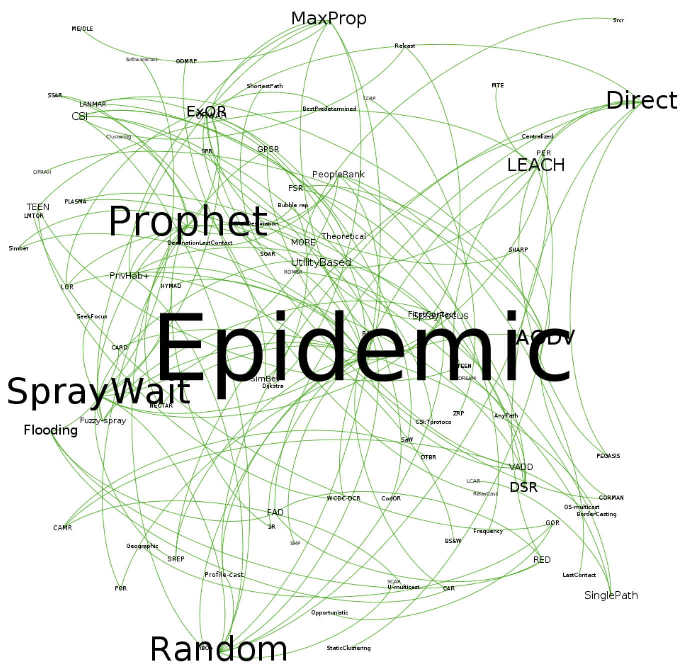


Figure 2.3.: Cloud graph of routing algorithms comparison, edges indicates a direct comparison between a pair of routing algorithms in a scientific article

Routing algorithms are vertices, while the edges are lines representing a comparison between routing algorithms connected by the line. In Figure 2.3, it can be seen that the Epidemic routing algorithm is the most compared.

2.5 OppNet simulation deployment

This section exhibits the main existing OppNets simulation tools. It also shows the elements and parameters that allow an OppNet simulation deployment and where those tools come from. Furthermore, this section reflects on how current simulation tools are used.

Section 2.4.1 briefly described Figure 2.1 which shows the elements involved when simulating an OppNet. The elements to deploy an OppNet simulation are input, output and software setup. The first element, the input, defines the network's behaviour, for example, nodes and message characteristics. The second element, the output, is the information obtained after the simulation, for example, routing algorithm performance metrics and delivery information of messages. Most of the time, performing a post-simulation analysis from the data obtained as the output may be necessary. Different simulation parameters are expected to return different outputs since the simulation is sensible to setup changes [35]. The final element, the software setup, is the configurations required to run the simulation software.

Several software alternatives allow simulating an OppNet, and most important, software that allows a simulation of routing algorithms behaviour. Among the software tools that allow simulating an OppNet are GloMoSim [36], OMNeT++ [37], DTN2 [38], HaggleSim [39], The ONE Simulator [40], ns-3 [41], Adyton [42] and MobEmu [43]. The simulation tools are mentioned in order of creation from 1998 to 2018. For OppNets research, the most used is The ONE simulator, reaching 62% of recent publications [12].

Furthermore, as is shown in Figure 2.1, scenarios are part of the inputs of a simulation deployment. The word scenarios are used indistinctly as a reference for a contact trace, mobility datasets, or even an entire simulation experiment. That is, there is no definition of a scenario in OppNets. Works like [26] define traces as datasets containing registers of nodes, and the information is either contacts, positions or both during a time. Simulation software is not strange to the use of contact or position traces. For example, The ONE Simulator [40] accepts trace datasets as input.

Traces are not always obtained from the real world and can be created computationally. These computationally created traces are known as synthetic traces. Other traces are a combination of the two types of traces, the real ones and the synthetic ones, known as hybrid traces. Real-world traces usually come from devices with GPS capabilities to obtain device positions during a period. For example, the well-known trace Asturias is a trace of the Asturias, Spain fire department, collected for one year. A total of 229 devices reported 19,462,339 locations with a granularity of 30 seconds [30].

However, a perception problem surrounds real-world traces; a cognitive bias. Although their real-world origin might have some characteristics, creating a trace suitable for simulation software might need some non-real-world assumptions, for example, that nodes have circular communication ranges [30] or inferred contacts [44]. In addition, their nature is random, which means that generalisation might be an issue. For instance, the difference among *the uiuc/uim dataset* [45], *the nus/contact dataset* [44] and *the unical/socialblueconn dataset* [46] might be challenging to assess. Because these traces were created within the contacts that students had in a university facility.

On the other hand, synthetic traces can be produced faster and might be as valuable as their real-world counterpart for evaluating routing algorithms [47]. Synthetic traces are a feasible solution to represent and deploy a simulation of an OppNet. With synthetic traces, OppNets can be represented because it is possible to select the traces that, as a whole, are a better representation of a desired OppNet environment due to their characteristics. The origin of the traces should not be an obstacle when considering using a trace. What is important is how well the trace represents an OppNet environment. Using several traces will be useless if they mimic the same network behaviour if the aim is to see how well a routing algorithm performs in different OppNets.

Nowadays, there is no creation methodology for new routing algorithms for OppNets. The following section reviews other research fields that evaluate the performance of algorithms.

2.6 Solutions from other fields

Evaluation and comparison have been used among researchers in other fields of study. This section reviews fair practices for performance evaluation and comparison of algorithms. Specifically, how the fields of data compression, linguistics and speech recognition handle the performance comparison problem when developing new algorithms.

This section introduces the term corpus that this thesis will use frequently, and it refers to a collection of representative data used to analyse the effectiveness of an algorithm's behaviour.

2.6.1 Data compression

Data compression aims to reduce the volume of the data preserving the quality, and it can be classified as either lossy or lossless compression. In lossy and lossless compression, the goal is to maintain quality by using the least amount of data to represent the information. As the name suggests, the original data can be obtained in lossless data compression, while in lossy compression, some information is lost.

Over the years, a large number of quite different compression algorithms have been proposed. Usually, when new compression algorithms are presented, some evaluation is provided by authors to demonstrate the superiority of their algorithm [48]. As well as routing algorithms, compression algorithms have problems showing the reliability of these evaluations, harming the repeatability of experiments and, thus, the scientific method.

One effective solution to provide fair comparison was using representative files shared across the scientific community in data compression. It is a collection of representative files to evaluate the effectiveness of the compression ratio[49]. The collection of representative files is called corpus. Calgary [50], Canterbury [48] and Silesia [51] are corpus used in lossless data compression.

Using a corpus to evaluate compression algorithms reduces bias and facilitates the experiments' reproducibility. The corpus should be in the public domain, freely distributed and with a distribution licence which allows wide usability

[52]. Furthermore, using corpus creates compression benchmarks, a standard compression ratio that other algorithms may be compared to. Nowadays, the criteria about the corpus are widely accepted in the compression field.

2.6.2 Linguistics

In linguistics, a corpus is a collection of texts or transcribed speeches that serve as a basis for linguistic analysis [53]. Authors in [54] define *"a corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research"*. It is a common belief that the use of a corpus began with the availability of computerised media. However, this is not the case in corpus-based linguistic studies. For example, biblical, lexicographical, dialect and grammatical corpus-based studies are pre-1960s examples [53].

After the pre-electronic corpus, in 1964, the first machine-readable corpus was presented, aimed to serve as a standard of comparison for studies focused on present-day English [55]. The texts in the corpus represent a broad range of stylistic aspects of written American English, setting a standard in corpus-based research.

The extensive use of corpus-based research in linguistics has provided some points of attention when using a corpus. (1) A corpus is like a photograph that captures the main features of a landscape. However, even with a careful design, not all local varieties can be represented, but the representatives [53]. (2) Representativeness and balance are significant issues because the language should be represented as a whole by the corpus, which means that the findings of corpus-based research must be generalised outside the corpus with a balance of the samples included within the corpus. (3) Even if a corpus is extensive and carefully designed, all the language characteristics could not be represented [54].

2.7 Speech recognition

Speech recognition aims to perform a recognition process transforming sounds into words. There are two main research fields, human speech recognition (HSR) and (ASR), sharing the same aim but their research approaches and objectives differ. HSR is interested in how humans, as listeners, recognise spoken utterances. Instead, ASR is focused on building algorithms that automatically recognise words in speech, limiting errors and under different environmental conditions [56].

A speech corpus can be used to train end-to-end recognition models in an unsupervised way [57]. The language for which the corpus was designed is not transferable to others. Of course, some corpora configuration allows working with multiple languages, such as *Euronews: a multilingual speech corpus for ASR*. Therefore, the intended use of a corpus is fundamental when choosing a corpus.

The state-of-the-art studies about routing algorithms are keen on a fair evaluation and comparison. The research fields of compression, linguistics, speech recognition or routing will not improve if applied to data that is not relevant or representative. Authors have been making efforts to provide fair performance evaluation and comparison. However, until today there is no commonly-agreed evaluation method. The use of self-selected traces, scenarios and setups hinders quality research. This thesis underlines the scientific community's efforts towards fair comparison in such challenged networks and saw no bad intention in their proposals. This thesis will introduce a novel general-purpose evaluation methodology suitable for OppNet routing algorithms in the following chapters. This novel general-purpose contribution is not linked to any specific routing algorithm. Instead, it is designed to provide a set of representative OppNet scenarios to evaluate any routing algorithm.

Part II

Proposal

3 Routing creation methodology

> *Measure that which is measurable and make measurable that which is not.*
>
> — **Galileo Galilei**

Nowadays, the exchange of information in networks occurs relatively quickly and efficiently. Sometimes it even goes unnoticed the existence of the fixed infrastructure behind them. Nevertheless, there is not always the availability of fixed network infrastructure, and the need to transmit information still prevails. Problems such as disruptions, delays and disconnections in fixed-infrastructure networks hinder the exchange of information. However, some networks can handle such problems so that nodes do not rely on fixed infrastructure but on peer communications. These networks are known as Opportunistic Networks (OppNets). Several specific labels could refer to those asynchronous communications within peers acting like carriers to exchange information to any network point. Still, in the end, these networks are under the OppNet umbrella.

In these challenged networks, the intelligence that allows the exchange of information is the routing algorithm. The routing algorithm decides whether a node should forward information or not, selecting a single or a group of peers that will carry the information. Those nodes repeat the process until the information reaches the intended node destination. Over the years, researchers in OppNets have created and proposed new routing algorithms, and until today, the scientific community have not agreed on a creation methodology. New routing algorithms usually arrive with a sort of evaluation process that validates their routing capabilities. Even so, obtaining an objective evaluation is difficult due to the particularities each research used to evaluate routing capabilities.

One proposal of this thesis is a methodology for creating routing algorithms that can improve objective quality, enabling reproducibility across results.

3.1 Motivation

In Chapter 2, non-traditional network communications were described. Those networks have connectivity problems, asymmetric bandwidths, delays and non-fixed topology. Intermittently connected devices create these networks taking advantage of their wireless capabilities. Such networks are known under the name of Opportunistic Networks (OppNets). The communication constraints of OppNets are due to the impossibility of deployment of a fixed infrastructure among wireless devices. Traditional protocols such as TCP/IP are unsuitable for providing communication solutions with such a constrained network. For this reason, routing algorithms such as Epidemic [14], Spray&Wait [16] and PRoPHET [15] have been created that adjust to specific communication challenges. In general terms, these routing algorithms implement a solution based on the store-carry-and-forward paradigm, which allows a message to be stored and transported until another suitable device(s) is found to forward the information.

The development of new routing algorithms is ongoing research. Over the years, researchers have been transforming routing ideas into routing algorithms. Research like [34] has studied the routing algorithms suitable for OppNets. Routing algorithms studied in [34] show different and interesting solutions to overcome the connectivity challenges. As depicted in [58], routing algorithms can be classified as direct transmission, flooding based, prediction based, coding based and context-based. The complexity of the routing problem seems to be well understood by the scientific community.

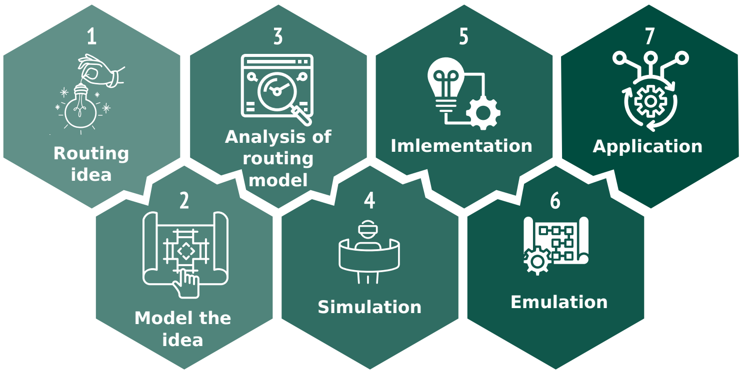


Figure 3.1.: Seven stage methodology for routing algorithm creation [59]

However, these different routing algorithms have been proposed separately with their assumptions, tools and definitions. This thesis proposes a step toward a methodology to create routing algorithms for OppNets, which could speed up the development of quality routing algorithms, enabling the capacity to reproduce and contrast routing performance among them. The following section introduces a complete creation methodology for developing routing algorithms.

3.2 Methodology to create new routing algorithms

As mentioned in Section 2, there is no methodology for creating routing algorithms. This thesis proposes a creation methodology based on seven stages as a path to transform a routing idea into an algorithm that can be implemented in the real world. Using a sensible, sound methodology is indispensable to get quality routing algorithms. The following methodology aims to observe the basic scientific method, allowing repeatability, fair comparison, and common scenario representation. This section proposes the basic steps of developing new routing algorithms.

3.2.1 Routing idea

The first stage, the routing idea, of this seven-stage creation methodology focuses on the idea of routing. This stage contains the creative part, where an outline of the main routing mechanism should be established. Generally, a routing algorithm has an essential functionality. For example, the Epidemic algorithm forwards messages to all the nodes whenever it establishes a contact. Instead, the algorithm Spray&Wait establishes in advance the number of copies.

In this stage, researchers could need to review other fields, settings, or state-of-the-art. Perhaps solutions from other fields can help strengthen a routing idea or even discard an unfeasible one.

3.2.2 Model

In the second stage, a model of the routing idea is created. A math-based conceptualisation could help to represent, explain and analyse the routing approach. Modelling a routing idea is establishing, among others, the entities involved, the parameters, the decision functions, the objective function and the restrictions. After defining those mentioned above, the routing behaviour can be modelled and analysed to validate the routing idea.

This stage clarifies the behaviour rules the routing algorithm will follow, and it might provide a creation development path since routing functions and components are atomised. Researchers will have an in-depth understanding of how the routing algorithm works.

3.2.3 Analysis

Once the routing idea is modelled as mentioned in Section 3.2.2, it can be analysed. During this analysis, the mechanisms and procedures can be checked, some theoretical results can be obtained, and fundamental limitations can be identified. Such results could help to make early-stage corrections before rushing into any of the stages described in the following sections. The analysis of the routing model should clarify operational functionalities that the routing algorithm will have.

3.2.4 Simulation

The next stage is simulation. The model can be tested in a given set of scenarios in a simulator. Even though these scenarios involve datasets that come from the real world (e.g., real traces from vehicles or people), or even if the simulator simulates very accurately all network protocols involved, the model under evaluation is usually executed based on pseudo-code. Results in this stage do not prove that the designed system can eventually be deployed and used for real use. Results obtained through simulation can be deceptive, creating a misleading feeling of scientific correctness. Indeed, as observed in [60], the credibility of simulation results tends to decrease as the use of simulation increases.

Moreover, a simulation is a trade-off between abstraction and representation. Abstraction is choosing which aspects of the real world should be simulated versus the ones that can be safely ignored [61]. Abstracting an OppNet to a simulation tool requires a careful examination and a deep understanding of the behaviours that the simulation will represent. Simulating a routing algorithm in all OppNets is not feasible due to computational processing limitations or the non-existence of a subset of them. Instead, simulating a routing algorithm in a set of representative OppNets will enable the scalability to perform in all OppNets.

3.2.5 Implementation

The final validation of a routing algorithm should always be based on real full-featured code (e.g. accounting for memory management or concurrency issues) rather than on the pseudo-code used in simulations. In this stage, a code is produced so the algorithm can be used in a real scenario, and the implementation itself shows the algorithm's feasibility.

3.2.6 Emulation

Testing real code in real conditions can be difficult and tricky, especially when these situations involve the mobility of hundreds of nodes for hundreds of hours. Emulation is an approach that helps with this regard, allowing to run real code in tightly controlled (and repeatable) conditions. This stage is the link between a proof-of-concept implementation and the deployment of software that is useful in the real world and behaves as predicted.

3.2.7 Application

The last stage of this methodology is testing the routing algorithm in a real environment with real devices and users. This is the ultimate test that shows how the designed algorithm behaves in the real world and allows evaluation.

3.3 Evaluation and comparison of new routing algorithms

Section 3.2 showed a methodology for the creation of new routing algorithms and Section 2.4 show the current approach for evaluation and comparison of routing algorithms. Among the seven-stages of the methodology, it is mentioned that evaluation is critical in the development process. An objective evaluation will tell if the routing algorithm acquires the expected performance, and a fair comparison will show the advances among the existing routing algorithms.

This section proposes the evaluation and comparison of routing algorithms to operate within the fourth and sixth stages of the seven-stage methodology. Doing so will contribute to enhancing the development. There are a few considerations for an objective evaluation and a fair comparison with other routing algorithms.

3.3.1 Metrics for performance evaluation

As mentioned in Section 2.3.2, metrics are used to evaluate the routing performance. Table 2.1 shows the diverse number of metrics found in literature. Delivery ratio, delivery delay and overhead are the most frequently used metrics. The delivery ratio computes the relationship between the number of messages created versus the number of messages delivered. The delivery delay is the difference in time that a message has to reach its destination. Overhead calculate the replication impact that a routing algorithm requires to deliver messages. These metrics are used in most of the contributions reviewed and listed in Table A.1.

The analysis of metrics is not within the scope of this thesis. In OppNets, metrics are standard functions that measure a property or process. Figure 3.2 depicts the metric taxonomy in OppNets, and it can be seen that metrics can be used to measure algorithm performance or routing performance.

3.4 OppNet scenarios

Section 2 shows that OppNet scenarios were mentioned in the literature for a while. Scenarios are essential in the creation methodology for routing algorithms proposed in Section 3.2, in specific in the stages of simulation and

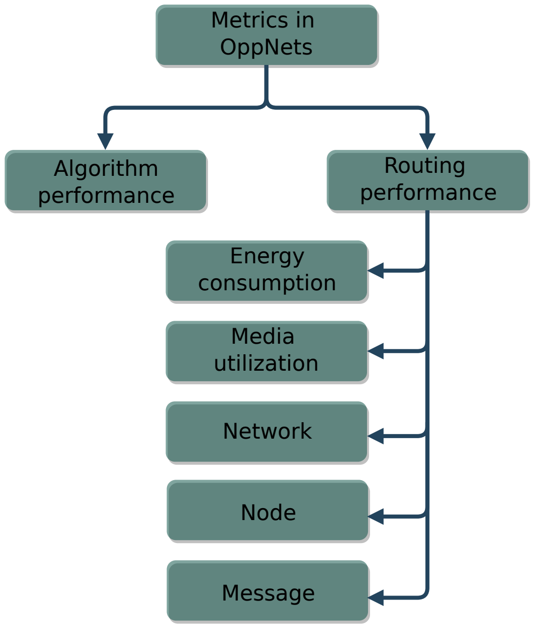


Figure 3.2.: Taxonomy distribution of metrics used in OppNets

emulation. Without scenarios that reproduce the behaviour of an OppNet, it would not be easy to simulate and emulate their behaviour. Until now, there is no clear definition of a scenario or its scope. This thesis proposes that scenarios should have a clear definition agreed upon by researchers, but more importantly, this thesis emphasises the necessity that the scenarios should be comparable between them. This last means, for example, that given two scenarios, researchers should be able to tell if those two scenarios are different or not.

By "different scenarios", this thesis refers to the network behaviour that a scenario resembles and its structure. A scenario has a structure and network behaviour. In other words, it is how the connections among nodes are established and the dynamics of how the information flows. The structure and network behaviour are cornerstones of this thesis because it is what matters when choosing one scenario instead of another. Until now, OppNet scenarios have not characterised their network structure or behaviour. An objective similarity evaluation among scenarios could be assessed by comparing the characteristics that describe scenarios.

This thesis proposes characterisation and definition of OppNet scenarios allowing the comparison between them. This proposal defines the elements involved and is not in the scope of the OppNet scenario definition.

Table 3.1.: Set of characteristics for a scenario definition, characteristics are classified as direct (D) and indirect (I).

Nº	Characteristic	Type	Description
1	Total number of nodes	D	$[\,nodes\,] \Rightarrow \{nodes \mid 192 < nodes < 960\}$
2	Nodes per group	D	$[\,nodes\_by\_group\,] \Rightarrow \{2^n \in \mathbb{Z} \mid 3 < n < 10\}$
3	Groups of nodes	D	$\{\ groups \in [1,2,3,4]\ \}$
4	Node's movements	D	$[\,movement\,] \Rightarrow \{movement \in [\ m_1,\ m_2,\ \ldots m_m]\}$
5	Node's speed	D	$1, 3, 7, 14$ *and* 27 *meters over second*
6	World size	D	$[width\ , height\,] \Rightarrow \{[width, height] \mid width, height \in [200 \ldots 3200]\}meters$
7	Area	D	$[\ area\,] \Rightarrow \{[area] \mid area \in [4000 \ldots 4160000]\}square\ meters$
8	Centrality	I	Measure of how much a given node is in-between other nodes
9	Inter-contact time	I	Time a node has no connection
10	Contact time	I	Duration time of the connection between two nodes
11	Contact time per minute	I	Contact-time within a minute window
12	Contact node ratio	I	Ratio of nodes contacted by a node
13	Popularity	I	Measure of the ratio of total unique connections
14	Window centrality	I	Mean centrality in a period
15	Encounters	I	Number of encounters
16	Sociability	I	Ratio of contacts
17	Total encounters	I	Total number of encounters within nodes

3.4.1 Scenario definition

As is explained in Section 2, nodes are the principal component of an OppNet scenario. Nowadays, the scenarios are considered a time-ordered list of contacts or positions that nodes have within the same OppNet. It is also mentioned in Section 2 that this information has been called contact traces. However, the contact traces also contain, in a non-explicit way, the corresponding network behaviour. Characterising a trace describes the intrinsic network behaviour of the trace with a vector of characteristics. In this thesis, *an OppNet scenario is denoted as a trace of positions characterised by a vector of characteristics*.

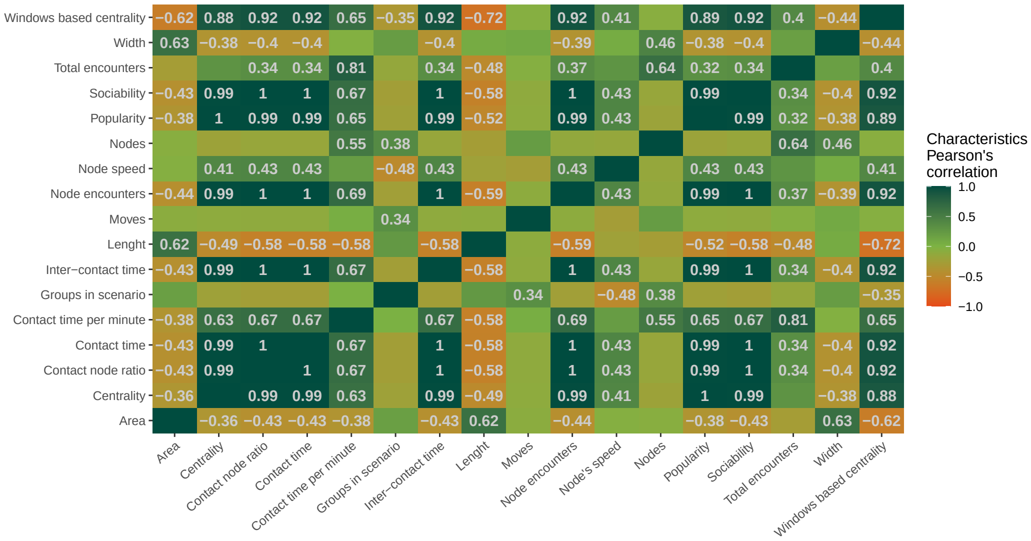


Figure 3.3.: Heat-map of Pearson correlation coefficients between scenarios' characteristics. Only significant Pearson's correlation coefficients are shown

3.4.2 Scenario characteristics

An OppNet scenario aims to imitate the network behaviour of an OppNet. Since OppNet scenarios represent reality, the number of characteristics can be extensive. However, not all characteristics from the real world are essential when representing an OppNet network behaviour. Table 3.1 shows the characteristics to describe an OppNet scenario.

The characteristics of an OppNet scenario shown in Table 3.1 can be classified as direct (D) and indirect (I). The direct characteristics of a scenario are the ones that can be quantified directly. For example, the number of nodes and the OppNet area are direct characteristics. On the other hand, indirect characteristics are the characteristics that are more complex to assess. Some of them might even require additional computing. For example, the number of contacts a node might have within a period cannot be calculated beforehand.

3.4.3 Characteristic correlation

The correlation study is a statistical measurement that represents the relationship between two characteristics, and this relationship can be expressed as a positive or negative correlation.

A positive correlation means that two characteristics vary in the same direction. If one increases, so does the other. A negative correlation occurs when the characteristics vary in an opposite direction. If one increases, the other decreases.

Figure 3.3 depicts a heat map to represent the Pearson correlation coefficients. As can be appreciated from Figure 3.3, some of the scenario characteristics have a high Pearson correlation. The high correlation might be due to the characteristics information source (connection-related information between nodes).

Figure 3.3 shows a clear correlation among scenario characteristics. Removing the highly correlated characteristics is one approach to dealing with them, selecting one of them as a representative characteristic of those correlated. However, removing characteristics that describe scenarios was considered a wrong approach because fewer scenario characteristics might hinder the scenario description accuracy. The high correlation helped to understand that selecting scenarios would not be straightforward and would require a backtracking process to achieve diversity and representativeness among scenarios in the corpus. The backtracking process uses additional information about the characteristics. Specifically, the variance of the characteristics was used in case there is a need to achieve representativeness and diversity objectives.

Centrality, Sociability, and Popularity characteristics are highly correlated with the Contact node ratio. The thesis studied some solutions to decide on the highly correlated characteristics. More information on the characteristics was obtained, for example, variance and standard deviation.

3.4.4 Scenario characterisation

The scenario characterisation is obtaining a vector of characteristics that describe an entire OppNet and its intrinsic network behaviour. The characteristics used are the ones depicted in Table 3.1. The characterisation of OppNet scenarios was accomplished in two phases.

The first phase is the obtaining of the scenario characteristics. In this stage, the characteristics that describe an OppNet are obtained. Those characteristics are shown in Table 3.1.

The second phase is to characterise an actual OppNet scenario. This phase obtains the values in each characteristic that describes the OppNet scenario, and the result is a vector of characteristics. This phase is a digest function that returns one vector of characteristics.

3.4.5 All-in-One scenario trap

Scenarios aim to represent OppNet environments, and a valuable scenario aims to be a good representation of reality. Representing an OppNet environment is not trivial because the representation must introduce as many elements as the real event contains. Representing a scenario with high-accuracy characteristics may not help network purposes because some characteristics might not influence how a network behaves. A high-accuracy representation could become an unrealistic task due to the number of characteristics involved. Our research found more than seventy characteristics used as tuning settings of scenarios. Given those seventy characteristics, even limiting each characteristic to a binary setup, the number of scenarios is unrealistic to manage.

Therefore, modelling a scenario requires a balance between simplification and real-world accuracy, which directly implies usability. Not every characteristic must be taken into account. Those characteristics that are not involved in the scenarios will be present in the other phases of the routing algorithm development. Oversimplification of characteristics could lead to a useless representation of the phenomena, and the results are not helpful. Our proposal claims that instead of building an "All-in-One" scenario, the development of a set of different scenarios, where those scenarios must give different performance results with the same algorithms.

3.5 Synthetic versus real-world traces

As mentioned in Chapter 2, real-world traces are often used in developing routing algorithms. It was also mentioned that whether the traces are a good representation of reality is scarce. Acquiring real-world data is challenging, time-consuming and even expensive. Having real-world traces is not a recipe for success in developing routing algorithms for OppNets.

This section studies the limitation of using real-world traces and how synthetic traces can be a suitable solution to overcome limitations such as representativeness, diversity and lack of information. First, the research conducted in this thesis has found that the representativeness of traces has not been studied. Datasets such as Asturias [30], Taxis Roma [31], Taxis San Francisco [32] and Cambridge/Haggle [33] are some of those often used. However, the selection of the real-world traces relies on the concept that the source of the traces is the collection from real-world and therefore they have an intrinsic value. This selection process introduces a cognitive bias.

Other problem with the real-world traces is the diversity, As is stated in Section 2.5, the difference among *the uiuc/UIM dataset* [45], *the nus/contact dataset* [44] and *the unical/socialblueconn dataset* [46] might be challenging to assess. Because these traces were created within the contacts that students had in a university facility. Real-world traces are source-dependent, meaning their origin will dictate the characteristics of such traces.

Last but not least, the lack of information on the traces is something that should not be overlooked. The information on the morphology of the traces, which in this thesis is described as the characterisation, is not always available. Synthetic traces mimics the behaviour of an OppNet and can fill the gap in information, representativeness and diversity. It can be created quicker than the real-world traces counterparts. Moreover, it is not only a matter of creation time. It is also a matter of the morphology of the trace. Synthetic traces is a flexible solution with scalability.

Table 3.2.: Indirect scenario characteristics measurement directives with references

N°	Characteristic	Measurement directive	Ref.
8	Centrality	Betweenness centrality computed as number of connections held by each node	[2]
9	Inter-contact time	Elapsed time each node has between contacts	[62]
10	Contact time	Elapsed time of the connection between two nodes	[62]
11	Contact time per minute	Contact-time within a period of one minute	[63]
12	Contact node ratio	Node contact ratio	[64]
13	Popularity	Unique peer-connections a node has	[65]
14	Window centrality	Centrality during a period	[66]
15	Encounters	Number of connections a node has	[67]
16	Sociability	Ratio of the number of contacts a node has to the total number of nodes	[68]
17	Total encounters	Summation of the number of connections within nodes	[67]

3.6 Measurement of indirect characteristics

Indirect characteristics are computed by processing the contact or position traces. The directives of the individual values, as well as the overall value, are depicted within the information contained in Table 3.1 and 3.2. For clarity, in Table 3.2 the number assigned to the characteristic corresponds to the number defined in Table 3.1. Except for the characteristic number seventeen, Total encounters, the characteristics shown in Table 3.2 are calculated in a two-step process. The first step is to calculate the characteristic individually in each node. Then, as a second and final step, the mean, variance and standard deviation values of the characteristics are calculated within the values of all or some nodes included in the scenario. For characteristic seventeen, the second step is the sum of the individual values of all nodes. The particular considerations are listed as follows:

- Centrality, Inter-contact time, Contact time, Contact node ratio and Encounters: mean of the individual measures of all nodes.
- Popularity and Sociability: mean of highest ten percentage measurements.
- Contact time per minute, Window betweenness centrality: mean of metrics within a period.
- Total encounters: accumulative measurement.

The next chapter of this thesis presents the main contribution of this thesis, which is the cornerstone in a path towards the standardisation of evaluation and comparison of routing algorithms in OppNet environments.

4 Corpus of scenarios for evaluation and comparison of routing algorithms

> *Users do not care about what is inside the box, as long as the box does what they need done.*
>
> — **Jef Raskin**
> about Human Computer Interfaces

Chapter 2 has shown that developing routing algorithms in OppNets can be improved using an algorithm creation methodology, particularly when comparing results. Although comparison is essential in research, scientific rigour cannot be assessed now when comparing the performance of routing algorithms. Section 2.6 of Chapter 2 also shows that a corpus helps in the algorithm development process proving to be a crucial part of the methodology. This section defines what an OppNet scenario is and how it can be characterised. Next, this section defines a complete methodology for the development of a corpus. Finally, this section presents a corpus for evaluating and comparing routing algorithms.

4.1 Corpus definition

A corpus, in the context of OppNets, is a collection of OppNet scenarios with two main features: first, all scenarios work together to cover all possible network behaviours, and second, the routing algorithms have different performance behaviour when routing messages in each scenario.

4.2 Quality requirements

The quality aspects taken into account for the corpus creation methodology are a line of action transversal to the sequence followed by the corpus creation methodology. The corpus aims to be a fair field for evaluating OppNet routing algorithms, providing a set of scenarios that can emulate real-world environments due to their characteristics. This thesis proposes a corpus creation methodology depicted in Figure 4.1 that will be explained in detail in Section 4.3. In addition, the corpus creation methodology presented in this research pursues the following requirements: coverage, scope, quality, and usability.

- **Coverage:** the coverage of the corpus should have representativeness for real-world environments, considering a significant difference between scenarios.
- **Scope:** the scope of the corpus should be the performance evaluation of routing algorithms in OppNets.
- **Quality:** the quality of each scenario of the corpus should be guaranteed by analysing the representativeness and diversity among other scenarios.
- **Usability:** corpus should be easy to use, and the scenarios should be adaptable to simulation software, where the evaluation of the performance of algorithms in OppNets is carried out.

4.3 Corpus creation methodology

This section presents the corpus creation methodology depicted in Figure 4.1. The corpus creation methodology has five well-delimited stages, each with specific inputs, outputs and tasks. The input information of one stage is the

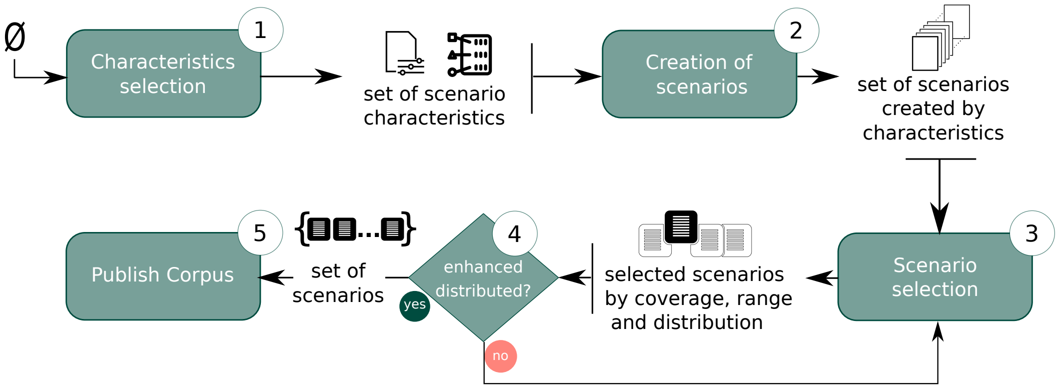


Figure 4.1.: Corpus creation methodology with backtracking stage for scenario selection assuring purpose, coverage, scope, quality, and usability requirements

output of the previous one, except for the first stage, which does not have an earlier stage.

The first stage, *characteristics selection*, decides those characteristics that describe a scenario. The selected seventeen characteristics are displayed in Table 3.1. A Pearson correlation [69] study of the selected characteristics was performed as shown in Figure 3.3. Some characteristics have a high correlation because are based in connections interaction between nodes. However, despite the redundancy and the high correlation, the characteristics reflect essential connectivity behaviours of the scenarios. This is why these high correlated characteristics remain within the selected characteristics.

Figure 3.3 made it clear that some characteristics were highly correlated. At first glance, one way to deal with high correlated characteristics is by removing them. However, removing characteristics that describe scenarios was considered a wrong approach because fewer scenario characteristics might hinder the scenario description accuracy. The high correlation helped to understand that selecting scenarios would not be straightforward and that it will require a backtracking process to achieve diversity and representativeness among scenarios in the corpus. The backtracking process uses additional information about the characteristics. Specifically, the variance of the characteristics was used in case there is a need to achieve representativeness and diversity objectives.

The second stage, *creation of scenarios*, received the characteristics found in stage one and then creates scenarios for the given characteristics. This stage generated over 200,000 OppNet scenarios, many of which had similar behaviour and therefore similar vector of characteristics. The scenarios with a similar vector of characteristics were considered equivalent.

The third and fourth stages, *scenario selection* and *enhanced distribution*, were loop connected. Each characteristic range was evenly divided into sub-ranges called windows. Then, a subset of scenarios was selected for each window, and this process sequentially loops through the list of characteristics. The number of scenarios was reduced because the scenarios should belong to all windows of the characteristics. If there was no scenario in the window, those empty-scenario windows were re-adjusted until scenarios were found.

When all the characteristics had been run through, and a representative number of scenarios had been obtained, stage four checks the diversity of the scenario collection. The loop is broken if the diversity of scenarios is fulfilled, which implies not having similar scenarios and that the distribution of characteristics manages to cover the entire range of each characteristic. Each scenario fulfils a part of the range of the characteristic. All scenarios, as a whole, complete the range of the characteristics.

The final stage, *publish corpus*, made the corpus of OppNet scenarios available for the research community. This assures the usability set as quality requirement shown in Section 4.2. The following section describes the corpus obtained following the corpus creation methodology presented in this section.

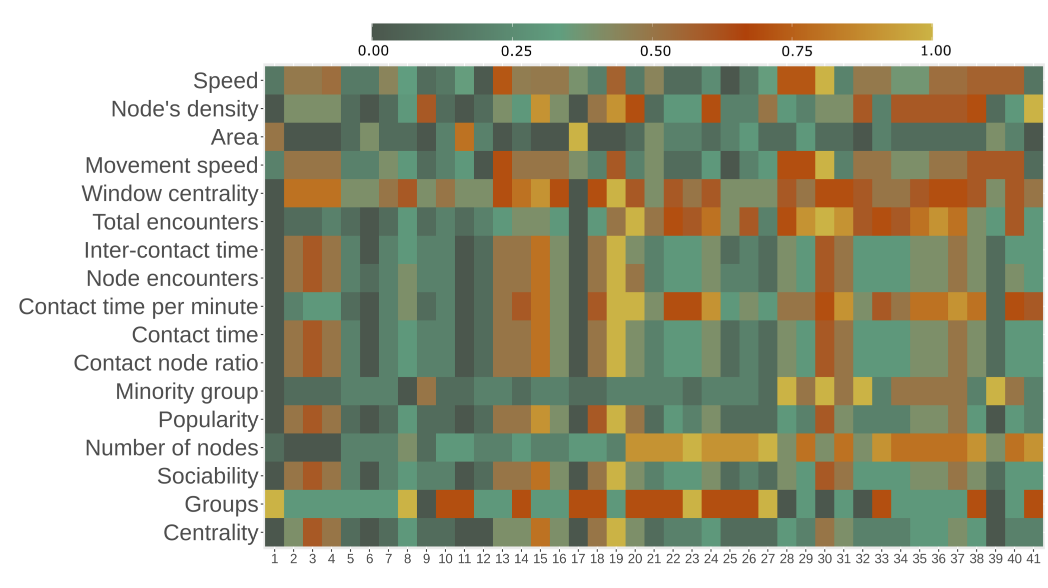
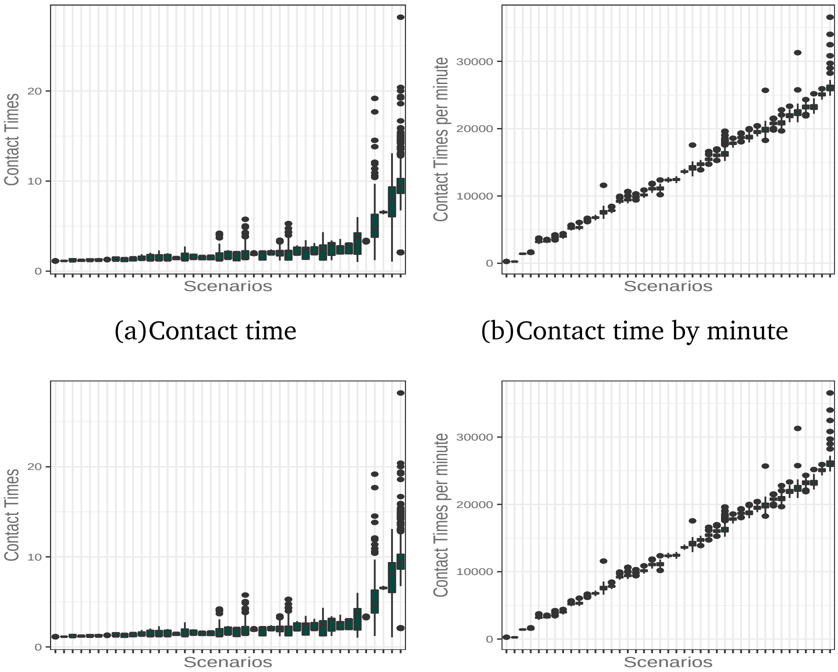


Figure 4.2.: Diversity representation of the forty-one scenario collection, which constitutes the first corpus for the performance evaluation of OppNet routing algorithms. The X-axis is the scenario's number, and the Y-axis represents the characteristics.

4.4 Corpus morphology

Section 3.2 describes creation of the corpus of OppNet scenarios which address the quality requirements mentioned in Section 4.2. Creating the corpus following the methodology returned forty-one scenarios with a balance between representativeness and diversity. The similarities among the scenarios increased with a number higher than forty-one, thus harming the diversity of the corpus. Moreover, some characteristics were not represented when the number was lower than forty-one. Therefore, the corpus is a collection of forty-one OppNet scenarios, and the characteristics and their distribution can be seen in Table 3.1 and Figure 4.3, respectively.

Scenarios in the corpus are identified with a number in the range [1-41]. Also, the corpus covers the range of each characteristic with the range of each scenario. In Figure 4.3, the X axis of each sub-figure represents the scenarios, and the Y axis represents the characteristic. Scenarios depicted in Figure 4.3 are not ordered by their number but by the value of the characteristic.



(a)Contact time (b)Contact time by minute

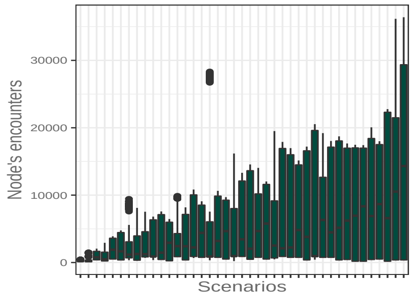
(c)Contact time (d)Contact time by minute

Figure 4.3.: Scenario characteristics range distribution (first four scenario characteristics out of eight)

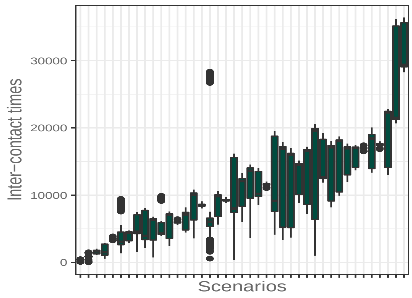
Furthermore, Figure 4.3 shows that node centrality, node inter-contact time and node sociability are characteristics with a high Pearson correlation among them. That is the reason why their figures have resemblance among them.

Figure 4.2 shows a study of the diversity of the corpus scenarios using a heatmap. It shows the relative intensity of characteristics of each of the scenarios in the corpus. Each column in Figure 4.2 is a scenario of the corpus. As it is mentioned throughout this thesis, the corpus will be expected to have representative as well as diverse scenarios. The Figure 4.2 shows that (1) there are no equal scenarios and (2) the distribution of the characteristics is uniform since there is no predominance of a single colour.

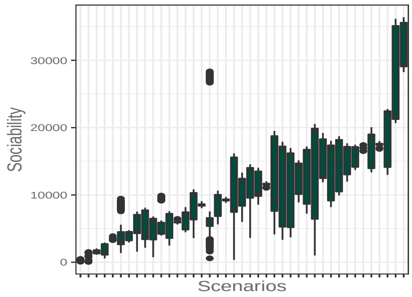
For usability reasons, each scenario of the corpus has two types of traces mentioned in Section 2.5, the contact traces and their homologous position traces. Furthermore, the granularity of the position traces is one second. Additionally, the contact traces can be obtained from their homologous based on the node positions but not the other way around.



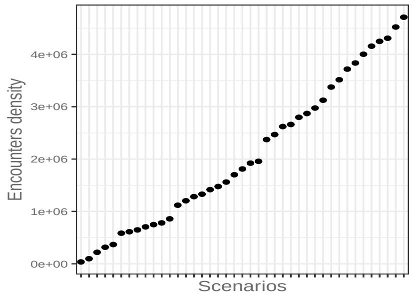
(d)Node encounters



(e)Inter-contact times



(f)Sociability



(g)Encounters density

Figure 4.3.: Scenario characteristics range distribution(last four scenario characteristics out of eight)

The scenarios simulate the speed of pedestrians, cyclists and two types of motorised vehicles. Those speeds are shown in Table 3.1. For this reason, up to four groups have been organised for each stage. Nodes among the same group share the speed and movement pattern. Movement patterns and node's speed are described in Table 3.1.

The number of nodes present in a scenario differs from one scenario to another. Still, the total number of nodes is distributed unevenly among the groups present in the scenarios with more than one group.

The morphology of the corpus depicted in Figure 4.3 is well distributed as a result of the methodologically selected scenarios. The following section assesses corpus behaviour when routing messages with a routing algorithms.

Section 4 explained the concept and characterisation of an OppNet scenario. It also defined and created a corpus to evaluate and compare OppNet routing algorithms. Section 4 also described the creation methodology and the morphology of the corpus obtained. The following section assesses the behaviour of the corpus.

4.5 Corpus performance appraisement

The experiment has been conducted over the Opportunistic Network Environment (The ONE simulator) [40] using the corpus of OppNet scenarios presented in Section 4.

Forty-one simulations were performed to assess the network representativeness of the corpus. In those simulations, node and message configurations are equal for all simulation setups. The forty-one simulations mean one simulation for each scenario of the corpus. The routing algorithm was the Epidemic, a routing algorithm replicating messages to every contacted node. The reason behind the selection of an epidemic routing algorithm for the experiment was the ability to flood the network with messages exhaustively. The Epidemic routing algorithm will forward a message to every node that it has contact with. Then, each recipient node will store the message until a new connection arises and repeat the forwarding process. An Epidemic routing algorithm will delete the message only when the assigned time to live of the message is reached.

As it was exposed in Section 2, routing algorithms aim to transmit messages from source to destination. For this reason, network behaviour could be expressed by how messages are delivered within the scenarios in the corpus. The simulations of the experiments have shown the behaviour of the corpus with the metrics related to message delivery. The metrics analysed were: the number of messages delivered, messages relayed, messages aborted, messages dropped, message hop-count, and the message buffer time.

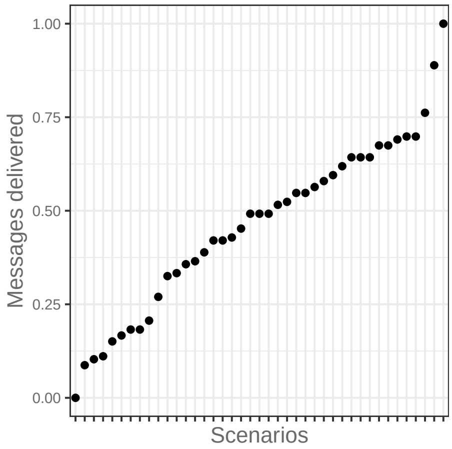
Figure 4.2 depicts the diversity within the characteristics vectors that define the scenarios in the corpus. In order to establish a difference among scenarios and, therefore, the corpus reliability, the scenario responses should be different between them. The response generated by each simulation has been analysed graphically to evidence their differences. Figure 4.4 presents the difference between the behaviours of the scenarios.

Figure 4.4 shows the difference of the response with eight sub-figures. Each sub-figure is a different metric. Axe $Scenarios$ in each sub-figure are the forty-one scenarios. Although all sub-figures contain the same scenarios, scenarios are not ordered equal from one sub-figure to another because they are ascending ordered by the metric that sub-figure represents. Axe Y in each sub-figure represents the normalised value of each scenario. Furthermore, each sub-figure depicts forty-one values in the [0-1] range since values are normalised.

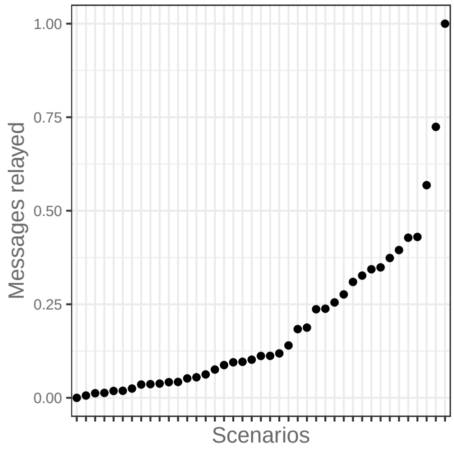
The results show a different response from one scenario to another, proving a different behaviour in each scenario. These results show the diversity among scenarios which is expressed in Section 4.2 as a corpus design requirement. Some areas are denser than others, but responses are well distributed overall.

4.6 Evaluation and comparison using the corpus

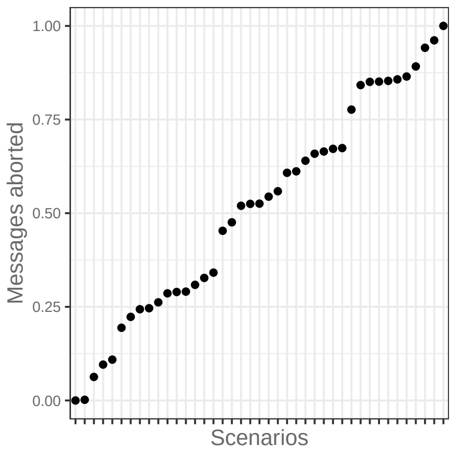
Now that the corpus contribution has been obtained via the methodology shown in Figure 3 and explained in Section 3.5, this section describes how the corpus can be used when a routing algorithm's evaluation and comparison process is needed. For the sake of clarity, some in-deep details are not included in this section, such as software configurations. The reader is asked to keep in mind that this section is intended to outline the usability of this study's main proposal rather than providing a closed recipe to use the corpus of OppNet scenarios.



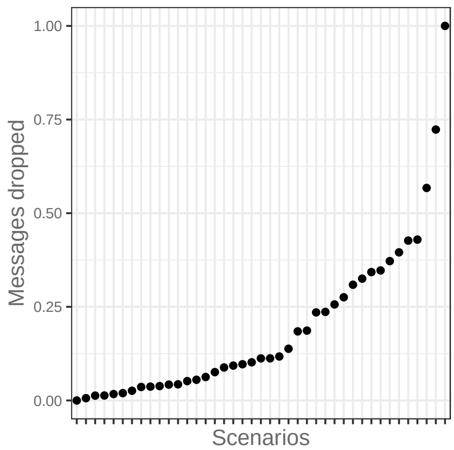
(a)Messages delivered



(b)Messages relayed

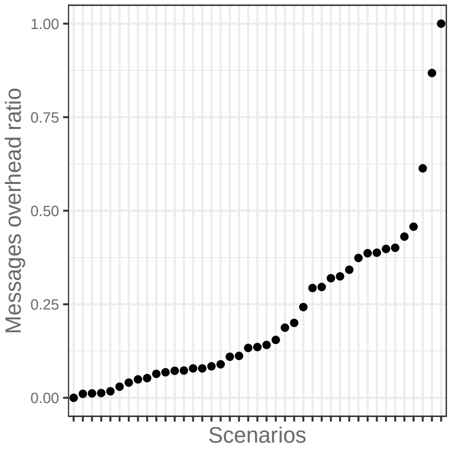


(c)Messages aborted

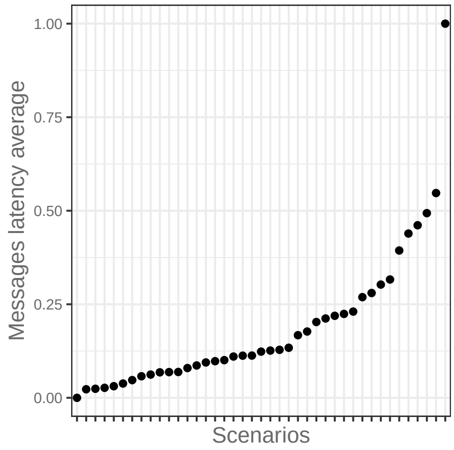


(d)Messages dropped

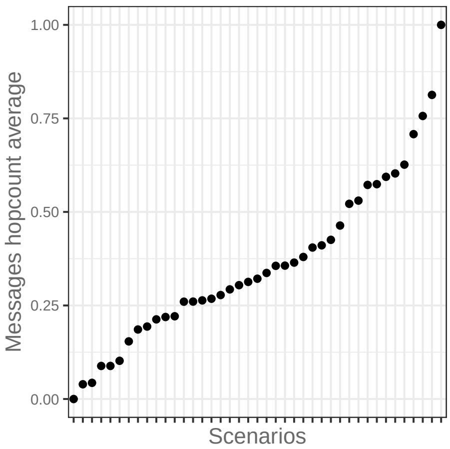
Figure 4.4.: Corpus scaled benchmarks for Epidemic routing, each dot in the figure is a scaled outcome ordered by the the metric, the identification number of scenarios is not showed and the order change among sub-figures.



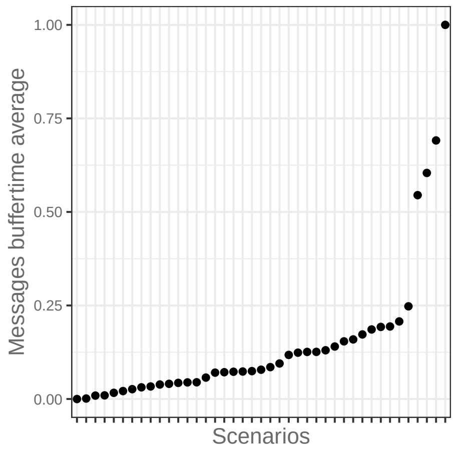
(e)Message overhead ratio



(f)Messages latency average



(g)Messages hopcount average



(h)Messages buffertime average

Figure 4.4.: Corpus scaled benchmarks for Epidemic routing, each dot in the figure is a scaled outcome ordered by the the metric, the identification number of scenarios is not showed and the order change among sub-figures.

When the research stage requires an evaluation of a routing algorithm, researchers interested in using a corpus will have to implement a simulation environment like the one shown in Figure 2.1. The authors should start by configuring the OppNet simulation software. After this, to use the corpus, researchers will have to download it. The corpus is available entirely free and without the necessity of login information. It is understood that in this stage, the routing algorithm that is going to be evaluated is already selected. Finally, authors might configure nodes and messages and establish the metrics that will retrieve the information to evaluate the performance. If the author desires to assess a comparison, the process will have to be repeated only by changing the routing algorithm. Then, the authors should compare the corresponding routing metrics obtained from the respective simulations.

This section has evaluated the behaviour of the corpus with a high replication algorithm to limit the response of the corpus when transmitting messages, and the results have shown that the corpus scenarios have different network behaviour between them. This result ratifies the positive assessment of a corpus. From now on, the scientific community has a collection of scenarios where their routing algorithms and features can be tested, thus avoiding scenario selection, reducing time and eliminating unintended bias. The corpus contributes towards establishing a proper bench-marking scheme for OppNet routing algorithms where the routing performance is not relative to other routing algorithms but overall.

5 Discussion

> *A lively discussion is usually helpful, because the hottest fire makes the hardest steel.*
>
> — **Tom Clancy**

Nowadays, OppNet routing algorithms can not be objectively evaluated neither compared because there is a lack of a globally accepted evaluation method. This situation hinders the development of new routing algorithms. The present proposal intends to contribute toward an objective evaluation methodology by providing an analytically selected collection of scenarios, a corpus. This proposal will help to ensure that evaluation results can be reliable, reproduced and contrasted in order to improve objective quality.

Researchers have tried to evaluate their proposals fairly, for example, by evaluating each other's proposals, using scenarios that other researchers have used, or selecting metrics that fit their proposals. However, these evaluation approaches have not overcome problems such as lack of reproducibility or inability to generalise routing algorithms to any scenario.

It is common practice in OppNets to use well-known scenarios with a clear intention of standardising evaluation methods. The problem, though, is not just a matter of using the same scenarios. If the routing algorithm being evaluated has to be general purposed, it is also a requirement that the scenarios being used are representative of all possible network situations. Therefore, any collection of scenarios is not the solution, and what is needed is a fine selection of representative scenarios.

Besides the existence of a representative corpus, it is as well important that it be used by the community. The corpus introduced in this study has been proved as representative by means of experimentation, and has been made publicly available.

This work is not intended to create a dilemma of whether or not the corpus should replace the well-known scenarios. Obtaining a simple corpus is not a difficult task. There are different ways of obtaining a collection of scenarios in a straightforward manner, for example using classical programming techniques such as random selection, trial and error, genetic algorithms, or even machine learning approaches. However, obtaining a representative corpus is complex and challenging. A representative one represents, as a whole, all-possible network behaviours. The selection of scenarios for a representative goes beyond a cherry-picking process, and each scenario is carefully analysed and compared with other scenarios. Still, the selection process may not matter as much as the corpus itself. The differences and representativeness of network behaviour that the corpus has is what determines if a corpus is useful or not.

The corpus presented in this work has been obtained via a creation methodology based on identifying the variables that characterise OppNet scenarios, methods to create OppNet scenarios and processes to assess differences and diversity among them. The differences and the representativeness of each scenario were carefully assessed. The results measured the representativeness and diversity of the corpus scenarios, showing significant differences. Therefore, it can be said that this is a representative corpus for objective evaluation. Having a representative corpus does not imply necessarily that it is the best. The scenarios of the corpus should be reviewed in the future, especially as new technologies emerge arising new network behaviours.

The corpus comprises simple scenarios where network behaviours are uniform. There might be environments where it is interesting to have non-uniform behaviours, for example, when defining strategies where the routing algorithm change depending on network conditions. These complex scenarios can be built, for instance, by concatenating simple scenarios from the corpus without unnecessarily expanding the number of scenarios in the corpus.

When there is a corpus, there is the risk of falling into the trap of developing tailored solutions that only work with the elements of this corpus. The behaviour of a routing algorithm should not be finely adjusted to have an outstanding performance in each corpus scenario since making a fine-tune would reduce the ability of a routing algorithm to extend the solution beyond the scenarios to the real world. Therefore, the routing model would not be able to generalise its routing abilities because the abilities would be too specific for the scenarios.

Another risk while developing routing algorithms for OppNets is exclusively focus, or paying too much attention, to simulation using the corpus. Simulation is just a part of the developing methodology, which should always be followed by an emulation stage, testing with actual implementations of the algorithms, and real world experimentation. Researchers should not overlook a complete methodology to convert a routing idea into real-world implementation.

Part III

Conclusions and Future Lines

6 Conclusion and Future Lines

In Chapter 1 the objectives of this thesis were presented: the main goal of this work was to provide a

6.1 Conclusions

From the state of the art, in the review of the methodologies for creating routing algorithms, it has been seen that, until now, there is no clear evidence to objectively evaluate, and thus compare the performance of these algorithms. Evaluating and comparing routing algorithms is a complex task, and the final quality of the algorithm significantly relies on it.

To right this wrong, this study proposed a potentially global-agreed for a fair evaluation and comparison of routing algorithms, a reference corpus of OppNet scenarios, which is a cornerstone in the design methodology. This corpus is a collection of forty-one methodologically obtained OppNet scenarios. These scenarios can be used to evaluate and thus compare the performance of routing algorithms. These scenarios were obtained using a creation procedure developed in this work which includes a backtracking process to enhance scenario diversity. This means that the corpus has the least number of scenarios, which, as a whole, represents most of the real-world OppNets.

Furthermore, for creating the corpus, it was necessary to characterise OppNets scenarios with a vector of characteristics. Such vectors are the basis for the analysis of similarities that lead to whether a scenario was a corpus member or not. The scenario is a node's contact trace described by a vector of seventeen characteristics.

The corpus presented can be an important tool to help researchers follow the scientific method, especially regarding reproducibility and standardisation aspects. These are essential features to improve quality research. The usefulness of the corpus requires that the community embraces it, using it for contrasting and evaluating routing performance results. The corpus is not static and should be revised to adapt to the needs; new technologies may require new scenarios in the future.

We look forward that this contribution will simplify and improve the development of routing algorithms in OppNets.

6.2 Future work

In the field of OppNets, ongoing research can be improved. This thesis has studied the OppNets and the current challenges. We also have acquired knowledge about our work, and the following section mentions the future investigation lines that are interesting to address based on this thesis outcomes.

6.2.1 On-the-fly scenario sensing

A projected line of research is the study of algorithm selection based on the sensing scenarios on-the-fly. This topic was discussed in Chapter 5 as an interesting topic. This topic involves two main challenges to address, non-uniform scenarios and routing algorithm selection. The first one, the non-uniform scenarios behaviours can be obtained with the corpus scenarios working as building blocks. The second challenge of this research topic, could lead to the study of routing interchangeability in OppNets, which means that a node will change the routing algorithm depending on the type of scenario that the mentioned node is real-time sensing.

6.2.2 Routing performance benchmarking scheme

This thesis has underlined the important aspects of a fair evaluation and comparison of routing algorithms. Evaluating and comparing the performance of a routing algorithm is a time-consuming task. The goals and benefits of doing this are not always obvious. Although the corpus is a vital evaluation tool, developing a routing performance benchmarking scheme will ensure that performance of one algorithm is not tied to a group of algorithms. With the implementation of a benchmarking scheme, routing performance will become an absolute result within the conditions of the benchmarking scheme.

6.2.3 Network behaviour layer

Last but not least, another interesting research topic is the implementation of an additional network information layer with network information. Such an approach could allow implementation similar to a traditional networking approach. The information layer of the network should be maintained by nodes that have bus behaviour. Those nodes in charge of keeping the network behaviour information updated should be able to sense and share the network behaviour information. The scenarios in the corpus are the network behaviour collection that can be used as the sensing source.

Part IV

Appendices

Appendices

A

> *Users do not care about what is inside the box, as long as the box does what they need done.*
>
> — **Jef Raskin**
> about Human Computer Interfaces

A.1 Routing algorithms

Table A.1.: Variables found in literature

Article	Metrics	Variables	Ref.
Epidemic Routing	Delivery Rate; Latency; Hops; Coverage floor; Dead message; Live message; Buffers; Lifetime dead messages	Radio range; message hop limit; buffer capacity	[14]
Comparing (DSR, AODV, DSDV i TORA)	Packet delivery ratio; Routing overhead; Path optimal	node movement speed; Pause time (mobility); number of sources	[70]
Dynamic Source Routing (DSR)	Optimal number transmissions		[71]
Temporally ordered routing algorithm (TORA)	Time complexity; Communication complexity		[72]
MaxProp	Delivery rate; median latency		[73]
Spray and wait	Average delivery delay; total number of transmissions; delivery ratio; maximum cluster size		[16]
Spray and focus	delivery delay; transmissions		[74]

Table A.1.: Variables found in literature

Article	Metrics	Variable	Ref.
Bubble rap	delivery ratio; delivery cost; hop-count distribution deliveries		[75]
Comparing (Direct Delivery and First contact, epidemic, spray&wait, prophet, fuzzy spray, SCAr, FAD, MaxProp, RAPID, Cluster-based routing, NECTAR, ORWAR, HiBOp, BubbleRap)	Delivery ratio; Overhead; Average delay		[76]
Comparing (AODV, OLSR, Epidemic, Spray&wait, prophet)	Delivery rate; Energy consumption; Delay		[77]
Prophet	delivery ratio; delivery delay; number of forwarded messages		[15]
Fuzzy-spray	delivery rate; delay; other the ONE metrics not in paper but mentioned	time; number of nodes	[78]
Survey	End-to-end delay; Packet loss; Location update time ; Location search time; data re-establishment time; Overhead		[79]
SCAR	delivery ration; message overhead; average number of hops per message	network size; buffer size; number of copies	[80]
FAD	Delivery ratio; average delay; average overhead		[81]
RAPID	minimum delay; average delay; delivery ratio; channel utilisation; Delivery ratio within deadline		[82]

Table A.1.: Variables found in literature

Article	Metrics	Variable	Ref.
cluster based routing	message delivery ratio; average message end-to-end delay; average number of control messages		[83]
NECTAR	delivery rate; hops per message; messages exchanged; delay; discarded messages		[17]
ORWAR	delivery rate; overhead; relative reduction of partial transmissions (RRPT)		[63]
HiBOp	buffer evolution (number of of messages); traffic overhead; message loss (percentage of); delay		[84]
Direct delivery	contact duration; inter-contact times; energy consumption; message delivery probability		[40]
PER	delivery ratio; end-to-end delivery latency		[85]
Simbet	message delivered; end-to-end delay; average number of hops per message; total number of forwards		[86]
SSAR	packet delivery ratio; total number of transmissions; selfishness satisfaction	Packet TTL; Packet Generation rate; Average number of social ties per node / willingness of nodes without social tie	[87]
SocialCast	message delivery ratio; network traffic; latency		[88]
HYMAD	delivery ratio; link and group stability		[89]
Timely-contact	delivery ratio; message transmissions; meta-data transmissions; message duplication; delay; storage usage		[90]
link-state	delivery rate / hop count; delivery rate; latency		[91]

Table A.1.: Variables found in literature

Article	Metrics	Variable	Ref.
Comparing (First contact, minimum expected delay, earliest delivery, earliest delivery with local queue, earliest delivery with all queue, linear program)	Fraction of data carried; delay; delivery ratio		[92]
CAR	delivery ratio; number of message; average delay		[93]
Peoplerank	success rate; delay		[94]
SREP	delivery ratio; delivery delay		[95]
3R	average delivery ratio; message overhead; average delivery time		[96]
Comparing (PRoPHET, Epidemic CGR)			[97]
Softwarecast	inter unique contact time; frequency; frequency (2); win ratio; latency; number of hops; win events		[98]
PrivHab+	latency; delivery ratio		[99]
Relcast	Latency; Node characteristic value ; Delivery ratio; Overhead		[100]
Social Profile-based Multicast (SPM)	Delivery ratio; Transmission cost; Average delay		[101]
RelayCast	Average delay; Average throughput per node (Kpbs)		[102]
Profile-cast	delivery ratio; delay; overhead		[103]
CSI	delivery ratio; delay; transmission overhead; storage overhead		[104]
CAMR	delivery ratio; average delay; data efficiency; overall efficiency		[105]

Table A.1.: Variables found in literature

Article	Metrics	Variable	Ref.
EBMR	delivery ratio; average delay; average hops; data efficiency; average number of queued messages		[106]
ODMRP	delivery ratio; control bytes transmitted / data byte delivered; total packets transmitted / data packet delivered		[107]
TZRP	query success ratio; route acquisition latency; total routing control overhead; pure proactive overhead; fuzzy proactive overhead; reactive overhead		[108]
CEDAR	Average path length; message complexity for route computation; time complexity for route computation; core usage; maximum available bandwith for computed routes; number of packets send; packets received; packets lost; packets rerouted; minimum delay		[109]
FSR	average overhead (bytes); avg throughput; normalised routing load; delay		[110]
DREAM	percentage of direct delivery; average delay		[111]
FSLS	throughput		[112]
CARD	reachable (percentage of); overhead		[113]
LANMAR	Delivery fraction; delay; normalised routing load; throughput		[114]
SHARP	packet overhead; loss rate; delay jitter		[115]
LEACH	energy dissipated; sensor lifetime; normalised energy dissipation	network diameter; percent of nodes that are cluster heads	[116]
PEGASIS		percentage of node death / initial energy	[117]

Table A.1.: Variables found in literature

Article	Metrics	Variable	Ref.
TEEN	Average energy dissipated; Total number of nodes alive		[118]
APTEEN	Average energy dissipated; Total number of nodes alive; Total number of data signals received at base station; Average Delay		[119]
WCDC-DCR	Average consumed energy; Network lifetime extension		[120]
ExOR	Packet delivery rate; Number of transmissions; number of uses	Distance; node pair; distance	[121]
OPRAH	Path length OPRAH / Path length AODV; length of OPRAH path	max speed	[122]
CCACK	throughput; number of total data transmissions; average fairness; throughput vs fairness; coding overhead; memory overhead ; header overhead	single flow / multiple flows; multiples flows	[123]
MORE	throughput; throughput per destination; average flow throughput; coding overhead; memory overhead; header overhead	number of flows	[124]
SOAR	average end-to-end goodput; fairness index	number of hops/ number of flows; number of flows	[125]
ROMER	throughput gain; successful transmissions	link failure prob / node failure prob	[126]
GOR	EOT; Path throughput	number of forwarding candidates / average number of neighbours per node	[127]
MGOR	end-to-end throughput bound	number if / maximum number of forwarding candidates	[128]
CodeOR	throughput ratio; data packet transmissions; throughput; total number of packets in network	number of hops on the shortest path	[129]

Table A.1.: Variables found in literature

Article	Metrics	Variable	Ref.
LCAR	transmission energy cost; route density average out-degree; route disconnections	Single-path distance; single-path distance; number of disconnected routes	[130]
PLASMA	throughput; delay	source rate	[131]
LOR	control overhead; End to end delay; Throughput	network diameter (hop count) / link quality coefficient; radio range distance / vehicle velocity / vehicle number; radio range distance / vehicle velocity / vehicle number	[132]
VADD	Delivery ratio; percentage of data packet dropped; delay; data traffic generated (per second / kbytes)	Data sending rate / data packet size	[133]
CORMAN	Delay; Delay jitter	Network dimension / node velocity	[134]
POR	delivery ratio; transmission delay	malicious nodes proportion	[135]

A.2 Synthetic traces generator

Algorithm 1: Partially-characterised synthetic trace generator for a given set of direct characteristics

Data: *Direct Characteristics*
Result: *Contact trace*
begin
end

Part V

Bibliography

Bibliography

[1] Helgason, Ólafur; Kouyoumdjieva, Sylvia T; Pajević, Ljubica; Yavuz, Emre A; Karlsson, Gunnar. A middleware for opportunistic content distribution. In: *Computer Networks* 107 (2016), pp. 178–193 (cit. on p. 9).

[2] Borrego, Carlos; Borrell, Joan; Robles, Sergi. Hey, influencer! Message delivery to social central nodes in social opportunistic networks. In: *Computer Communications* 137 (2019), pp. 81–91 (cit. on pp. 9, 10, 40).

[3] Chen, Depeng; Borrego, Carlos; Navarro-Arribas, Guillermo. A Privacy-Preserving Routing Protocol Using Mix Networks in Opportunistic Networks. In: *Electronics* 9.11 (2020), pp. 2–15 (cit. on p. 9).

[4] Danielis, Peter; Karlsson, Gunnar. Survey of mobile opportunistic networks for parallel data dissemination and processing. In: *KuVS-Fachgesp* 1 (2020), pp. 1–3 (cit. on p. 10).

[5] Sarros, Christos-Alexandros; Demiroglou, Vassilis; Tsaoussidis, Vassilis. Intermittently-connected IoT devices: Experiments with an NDN-DTN architecture. In: *18th Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, USA*. 9 January 2021 (cit. on p. 10).

[6] Nayyar, Anand; Batth, Ranbir Singh; Ha, Dac Binh; Sussendran, G. Opportunistic networks: present scenario-a mirror review. In: *International Journal of Communication Networks and Information Security* 10.1 (2018), pp. 223–241 (cit. on p. 10).

[7] Borrego, Carlos; Castillo, Sergio; Robles, Sergi. Striving for sensing: Taming your mobile code to share a robot sensor network. In: *Information Sciences* 277 (2014), pp. 338–357 (cit. on p. 10).

[8] Trifunovic, Sacha; Kouyoumdjieva, Sylvia T.; Distl, Bernhard, et al. A decade of research in opportunistic networks: Challenges, relevance, and future directions. In: *IEEE Communications Magazine* 55.1 (2017), pp. 168–173 (cit. on p. 10).

[9] Du, Zhaoyang; Wu, Celimuge; Chen, Xianfu, et al. A VDTN scheme with enhanced buffer management. In: *Wireless Networks* 26.3 (2020), pp. 1537–1548 (cit. on p. 10).

[10] Borrego, Carlos; Amadeo, Marica; Molinaro, Antonella, et al. Forwarding in opportunistic information-centric networks: an optimal stopping approach. In: *IEEE Communications Magazine* 58.5 (2020), pp. 56–61 (cit. on p. 10).

[11] Cho, Hsin-Hung; Chen, Chi-Yuan; Shih, Timothy K; Chao, Han-Chieh. Survey on underwater delay/disruption tolerant wireless sensor network routing. In: *IET Wireless Sensor Systems* 4.3 (2014), pp. 112–121 (cit. on p. 10).

[12] Kuppusamy, Vishnupriya; Thanthrige, Udaya Miriya; Udugama, Asanga; Förster, Anna. Evaluating forwarding protocols in opportunistic networks: Trends, advances, challenges and best practices. In: *Future Internet* 11.5 (2019), p. 113 (cit. on pp. 11, 12, 18).

[13] Alajeely, Majeed; Doss, Robin; Ahmad, Asma'a. Routing protocols in opportunistic networks–a survey. In: *IETE Technical Review* 35.4 (2018), pp. 369–387 (cit. on pp. 11, 15).

[14] Vahdat, Amin; Becker, David. *Epidemic routing for partially connected ad hoc networks*. 2000 (cit. on pp. 11, 28, 67).

[15] Lindgren, Anders; Doria, Avri; Schelen, Olov. Probabilistic routing in intermittently connected networks. In: *International Workshop on Service Assurance with Partial and Intermittent Resources, Fortaleza, Brazil*. 6 August 2004 (cit. on pp. 11, 28, 68).

[16] Spyropoulos, Thrasyvoulos; Psounis, Konstantinos; Raghavendra, Cauligi S. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In: *SIGCOMM05: ACM SIGCOMM 2005 Conference, Pennsylvania, USA*. 26 August 2005 (cit. on pp. 11, 28, 67).

[17] De Oliveira, Etienne CR; De Albuquerque, Celio VN. NECTAR: a DTN routing protocol based on neighborhood contact history. In: *SAC09: The 2009 ACM Symposium on Applied Computing, Honolulu, Hawaii*. 8March 2009 (cit. on pp. 11, 69).

[18] Grasic, Samo; Lindgren, Anders. Revisiting a remote village scenario and its DTN routing objective. In: *Computer Communications* 48 (2014), pp. 133–140 (cit. on p. 11).

[19] Kaner, Cem; Bond, Walter P. Software engineering metrics: What do they measure and how do we know. In: *In 10th International Software Metrics Symposium, Chicago, USA*. 11 September 2004 (cit. on p. 12).

[20] Grasic, Samo; Lindgren, Anders. An Analysis of Evaluation Practices for DTN Routing Protocols. In: *Seventh ACM International Workshop on Challenged Networks, Istanbul, Turkey*. 22 August 2012 (cit. on p. 12).

[21] Petz, Agoston; Enderle, Justin; Julien, Christine. A framework for evaluating dtn mobility models. In: *2nd International Conference on Simulation Tools and Techniques, Rome, Italy*. 6 March 2009 (cit. on p. 12).

[22] Sandulescu, Gabriel. Resource-Aware routing in delay and disruption tolerant networks. PhD thesis. University of Luxembourg, Luxembourg, 2011 (cit. on p. 12).

[23] Angius, Fabio; Gerla, Mario; Pau, Giovanni. Bloogo: Bloom filter based gossip algorithm for wireless NDN. In: *ACM workshop on Emerging Name-Oriented Mobile Networking Design-Architecture, Algorithms, and Applications, Carolina, USA*. 11 June 2012 (cit. on p. 12).

[24] Zhang, Yu Jin. Evaluation and comparison of different segmentation algorithms. In: *Pattern recognition letters* 18.10 (1997), pp. 963–974 (cit. on p. 13).

[25] Dede, Jens; Förster, Anna; Hernández-Orallo, Enrique, et al. Simulating Opportunistic Networks: Survey and Future Directions. In: *IEEE Communications Surveys and Tutorials* 20.2 (2018), pp. 1547–1573. arXiv: `1712.01905` (cit. on pp. 14, 15).

[26] Kotz, David; Henderson, Tristan; Abyzov, Ilya; Yeo, Jihwang. *CRAWDAD dataset dartmouth/campus (v. 2009-09-09)*. Downloaded from `https://crawdad.org/dartmouth/campus/20090909`. Sept. 2009 (cit. on pp. 15, 18).

[27] Manfredi, Victoria; Crovella, Mark; Kurose, Jim. Understanding stateful vs stateless communication strategies for ad hoc networks. In: *17th annual international conference on Mobile computing and networking, Las Vegas, USA*. 19 September 2011 (cit. on p. 15).

[28] Souza, Camilo; Mota, Edjair; Manzoni, Pietro, et al. Friendly-drop: A social-based buffer management algorithm for opportunistic networks. In: *2018 Wireless Days (WD), Dubai, United Arab Emirates*. 3 April 2018 (cit. on p. 15).

[29] Borrego, Carlos; Borrell, Joan; Robles, Sergi. Efficient broadcast in opportunistic networks using optimal stopping theory. In: *Ad Hoc Networks* 88 (2019), pp. 5–17 (cit. on p. 15).

[30] Cabrero, Sergio; Garcia, Roberto; García, Xabiel G.; Melendi, David. *CRAWDAD dataset oviedo/asturies-er (v. 2016-08-08)*. Downloaded from `https://crawdad.org/oviedo/asturies-er/20160808`. Aug. 2016 (cit. on pp. 15, 18, 19, 39).

[31] Bracciale, Lorenzo; Bonola, Marco; Loreti, Pierpaolo, et al. *CRAWDAD dataset roma/taxi (v. 2014-07-17)*. Downloaded from `https://crawdad.org/roma/taxi/20140717`. July 2014 (cit. on pp. 15, 39).

[32] Piorkowski, Michal; Sarafijanovic-Djukic, Natasa; Grossglauser, Matthias. *CRAWDAD dataset epfl/mobility (v. 2009-02-24)*. Downloaded from `https://crawdad.org/epfl/mobility/20090224`. Feb. 2009 (cit. on pp. 15, 39).

[33] Akestoridis, Dimitrios-Georgios. *CRAWDAD dataset uoi/haggle (v. 2016-08-28): derived from cambridge/haggle (v. 2009-05-29)*. Downloaded from `https://crawdad.org/uoi/haggle/20160828/one`. Aug. 2016 (cit. on pp. 15, 39).

[34] Kuppusamy, Vishnupriya; Thanthrige, Udaya Miriya; Udugama, Asanga; Förster, Anna. Evaluating forwarding protocols in opportunistic networks: trends, advances, challenges and best practices. In: *Future Internet* 11.5 (2019), p. 113 (cit. on pp. 16, 28).

[35] Abdelkader, Tamer; Naik, Kshirasagar; Nayak, Amiya; Goel, Nishith; Srivastava, Vineet. A performance comparison of delay-tolerant network routing protocols. In: *IEEE Network* 30.2 (2016), pp. 46–53 (cit. on p. 18).

[36] Bajaj, Lokesh; Takai, Mineo; Ahuja, Rajat, et al. Glomosim: A scalable network simulation environment. In: *UCLA computer science department* Technical report 990027 (1999), pp. 1–12 (cit. on p. 18).

[37] Varga, Andras. OMNeT++. In: *Modeling and tools for network simulation*. Ed. by Frederiksen, Norman O.; Gulliksen, Harold. Springer: Berlin, Germany, 2010, pp. 35–59 (cit. on p. 18).

[38] Fall, Kevin; Ott, J. *Delay-Tolerant Networking Research Group-DTNRG*. 2002 (cit. on p. 18).

[39] Su, Jing; Scott, James; Hui, Pan, et al. Haggle: Seamless networking for mobile applications. In: *International Conference on Ubiquitous Computing, Innsbruck, Austria*. 16 September 2007 (cit. on p. 18).

[40] Keränen, Ari; Ott, Jörg; Kärkkäinen, Teemu. The ONE Simulator for DTN Protocol Evaluation. In: *2nd International Conference on Simulation Tools and Techniques, Rome, Italy*. 2 March 2009 (cit. on pp. 18, 50, 69).

[41] Riley, George F; Henderson, Thomas R. The ns-3 network simulator. In: *Modeling and tools for network simulation*. Ed. by Frederiksen, Norman O.; Gulliksen, Harold. Springer: Berlin, Germany, 2010, pp. 15–34 (cit. on p. 18).

[42] Papanikos, Nikolaos; Akestoridis, Dimitrios-Georgios; Papapetrou, Evangelos. *CRAWDAD toolset tools/simulate/uoi/adyton (v. 2016-04-21)*. Downloaded from `https://crawdad.org/tools/simulate/uoi/adyton/20160421`. Apr. 2016 (cit. on p. 18).

[43] Ciobanu, Radu-Ioan; Marin, Radu-Corneliu; Dobre, Ciprian. Mobemu: a framework to support decentralized ad-hoc networking. In: *Modeling and Simulation in HPC and Cloud Systems*. Ed. by Joanna Kołodziej Florin Pop, Ciprian Dobre. Springer: Berlin, Germany, 2018, pp. 87–119 (cit. on p. 18).

[44] Srinivasan, Vikram; Motani, Mehul; Ooi, Wei Tsang. *CRAWDAD dataset nus/contact (v. 2006-08-01)*. Downloaded from `https://crawdad.org/nus/contact/20060801`. Aug. 2006 (cit. on pp. 19, 39).

[45] Nahrstedt, Klara; Vu, Long. *CRAWDAD dataset uiuc/uim (v. 2012-01-24)*. Downloaded from `https://crawdad.org/uiuc/uim/20120124`. Jan. 2012 (cit. on pp. 19, 39).

[46] Caputo, Antonio; Socievole, Annalisa; Rango, Floriano De. *CRAWDAD dataset unical/socialblueconn (v. 2015-02-08)*. Downloaded from `https://crawdad.org/unical/socialblueconn/20150208`. Feb. 2015 (cit. on pp. 19, 39).

[47] Thiebaut, Dominique; Wolf, Joel L.; Stone, Harold S. Synthetic traces for trace-driven simulation of cache memories. In: *IEEE Transactions on computers* 41.04 (1992), pp. 388–410 (cit. on p. 19).

[48] Arnold, Ross; Bell, Tim. A corpus for the evaluation of lossless compression algorithms. In: *DCC'97. Data Compression Conference, Snowbird, USA*. 25 March 1997 (cit. on p. 20).

[49] Islam, Md Rafiqul; Rajon, SA Ahsan. On the design of an effective corpus for evaluation of Bengali Text Compression Schemes. In: *2008 11th International Conference on Computer and Information Technology, Khulna, Bangladesh*. 27 December 2008 (cit. on p. 20).

[50] Usama, Muhammad; Malluhi, Qutaibah M; Zakaria, Nordin; Razzak, Imran; Iqbal, Waheed. An efficient secure data compression technique based on chaos and adaptive Huffman coding. In: *Peer-to-Peer Networking and Applications* 14.5 (2021), pp. 2651–2664 (cit. on p. 20).

[51] Gupta, Apoorv; Bansal, Aman; Khanduja, Vidhi. Modern lossless compression techniques: Review, comparison and analysis. In: *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*. IEEE. 2017, pp. 1–8 (cit. on p. 20).

[52] Holub, Jan; Reznicek, Jakub; Šimek, Filip. Lossless data compression testbed: Excom and prague corpus. In: *2011 Data Compression Conference*. IEEE. 2011, pp. 457–457 (cit. on p. 21).

[53] Kennedy, Graeme. *An introduction to corpus linguistics*. Routledge, 2014 (cit. on p. 21).

[54] De Pauw, Guy. *Developing Linguistic Corpora—A Guide to Good Practice Martin Wynne (ed.)*. 2006 (cit. on p. 21).

[55] Francis, W Nelson; Kucera, Henry. Brown corpus manual. In: *Letters to the Editor* 5.2 (1979), p. 7 (cit. on p. 21).

[56] Scharenborg, Odette. Reaching over the gap: A review of efforts to link human and automatic speech recognition research. In: *Speech Communication* 49.5 (2007), pp. 336–347 (cit. on p. 22).

[57] Gretter, Roberto. Euronews: a multilingual speech corpus for ASR. In: *LREC*. 2014, pp. 2635–2638 (cit. on p. 22).

[58] Poonguzharselvi, B; Vetriselvi, V. Survey on routing algorithms in opportunistic networks. In: *2013 International Conference on Computer Communication and Informatics*. IEEE. 2013, pp. 1–5 (cit. on p. 28).

[59] Freire, Diego; Robles, Sergi; Borrego, Carlos. Towards a Methodology for the Development of Routing Algorithms in Opportunistic Networks. In: *The Sixteenth International Conference on Wireless and Mobile Communications ICWMC 2020, Oporto, Portugal*. 19 October 2020 (cit. on p. 29).

[60] Kurkowski, Stuart; Camp, Tracy; Colagrosso, Michael. MANET simulation studies: the incredibles. In: *ACM SIGMOBILE Mobile Computing and Communications Review* 9.4 (2005), pp. 50–61 (cit. on p. 31).

[61] Swarup, Samarth. Adequacy: what makes a simulation good enough? In: *2019 Spring Simulation Conference (SpringSim)*. IEEE. 2019, pp. 1–12 (cit. on p. 31).

[62] Karamshuk, Dmytro; Boldrini, Chiara; Conti, Marco; Passarella, Andrea. Human mobility models for opportunistic networks. In: *IEEE Communications Magazine* 49.12 (2011), pp. 157–165 (cit. on p. 40).

[63] Sandulescu, Gabriel; Nadjm-Tehrani, Simin. Opportunistic DTN routing with window-aware adaptive replication. In: *Proceedings of the 4th Asian Conference on Internet Engineering*. 2008, pp. 103–112 (cit. on pp. 40, 69).

[64] Yuan, Peiyan; Wang, Chenyang. OPPO: An optimal copy allocation scheme in mobile opportunistic networks. In: *Peer-to-Peer Networking and Applications* 11.1 (2018), pp. 102–109 (cit. on p. 40).

[65] Schurgot, Mary R; Comaniciu, Cristina; Jaffres-Runser, Katia. Beyond traditional DTN routing: social networks for opportunistic communication. In: *IEEE Communications Magazine* 50.7 (2012), pp. 155–162 (cit. on p. 40).

[66] Settawatcharawanit, Tossaphol; Yamada, Shigeki; Haque, Md Enamul; Rojviboonchai, Kultida. Message dropping policy in congested social delay tolerant networks. In: *The 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. IEEE. 2013, pp. 116–120 (cit. on p. 40).

[67] Bhattacharjee, Suman; Roy, Siuli; Ghosh, Sukumar; DasBit, Sipra. Exploring the impact of connectivity on dissemination of post disaster situational data over DTN. In: *Proceedings of the 18th International Conference on Distributed Computing and Networking*. 2017, pp. 1–4 (cit. on p. 40).

[68] Boldrini, Chiara; Conti, Marco; Passarella, Andrea. Social-based autonomic routing in opportunistic networks. In: *Autonomic Communication*. Springer, 2009, pp. 31–67 (cit. on p. 40).

[69] Freedman, David; Pisani, Robert; Purves, Roger. *Statistics (international student edition)*. 4th. New York, USA: WW Norton & Company, 2007 (cit. on p. 45).

[70] Broch, Josh; Maltz, David A; Johnson, David B; Hu, Yih-Chun; Jetcheva, Jorjeta. A performance comparison of multi-hop wireless ad hoc network routing protocols. In: *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*. 1998, pp. 85–97 (cit. on p. 67).

[71] Johnson, David B; Maltz, David A. Dynamic source routing in ad hoc wireless networks. In: *Mobile computing*. Springer, 1996, pp. 153–181 (cit. on p. 67).

[72] Park, Vincent Douglas; Corson, M Scott. A highly adaptive distributed routing algorithm for mobile wireless networks. In: *Proceedings of INFOCOM'97*. Vol. 3. IEEE. 1997, pp. 1405–1413 (cit. on p. 67).

[73] Burgess, John; Gallagher, Brian; Jensen, David D; Levine, Brian Neil, et al. MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks. In: *Infocom*. Vol. 6. Barcelona, Spain. 2006 (cit. on p. 67).

[74] Spyropoulos, Thrasyvoulos; Psounis, Konstantinos; Raghavendra, Cauligi S. Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility. In: *Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerComW'07)*. IEEE. 2007, pp. 79–85 (cit. on p. 67).

[75] Hui, Pan; Crowcroft, Jon; Yoneki, Eiko. Bubble rap: Social-based forwarding in delay-tolerant networks. In: *IEEE transactions on mobile computing* 10.11 (2010), pp. 1576–1589 (cit. on p. 68).

[76] Massri, Khalil; Vernata, Alessandro; Vitaletti, Andrea. Routing protocols for delay tolerant networks: a quantitative evaluation. In: *Proceedings of the 7th ACM workshop on Performance monitoring and measurement of heterogeneous wireless and wired networks*. 2012, pp. 107–114 (cit. on p. 68).

[77] Almeida, Virgil Del Duca; Oliveira, André B; Macedo, Daniel F; Nogueira, José Marcos S. Performance evaluation of manet and dtn routing protocols. In: *2012 IFIP Wireless Days*. IEEE. 2012, pp. 1–6 (cit. on p. 68).

[78] Mathurapoj, Akadet; Pornavalai, Chotipat; Chakraborty, Goutam. Fuzzy-spray: Efficient routing in delay tolerant ad-hoc network based on fuzzy decision mechanism. In: *2009 IEEE International Conference on Fuzzy Systems*. IEEE. 2009, pp. 104–109 (cit. on p. 68).

[79] Moreira, Waldir; Mendes, Paulo. Survey on opportunistic routing for delay/disruption tolerant networks. In: *Dept. Inform. Syst. Technol., Univ. Lusófona, Lisbon, Portugal, Tech. Rep. SITI-TR-11-02* (2011) (cit. on p. 68).

[80] Pásztor, Bence; Musolesi, Mirco; Mascolo, Cecilia. Opportunistic mobile sensor data collection with scar. In: *2007 IEEE International Conference on Mobile Adhoc and Sensor Systems*. IEEE. 2007, pp. 1–12 (cit. on p. 68).

[81] Wang, Yu; Wu, Hongyi. Delay/fault-tolerant mobile sensor network (dft-msn): A new paradigm for pervasive information gathering. In: *IEEE Transactions on mobile computing* 6.9 (2007), pp. 1021–1034 (cit. on p. 68).

[82] Balasubramanian, Aruna; Levine, Brian; Venkataramani, Arun. DTN routing as a resource allocation problem. In: *Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*. 2007, pp. 373–384 (cit. on p. 68).

[83] Dang, Ha; Wu, Hongyi. Clustering and cluster-based routing protocol for delay-tolerant mobile networks. In: *IEEE Transactions on Wireless Communications* 9.6 (2010), pp. 1874–1881 (cit. on p. 69).

[84] Boldrini, Chiara; Conti, Marco; Jacopini, Jacopo; Passarella, Andrea. Hibop: a history based routing protocol for opportunistic networks. In: *2007 IEEE international Symposium on a world of wireless, mobile and multimedia networks*. IEEE. 2007, pp. 1–12 (cit. on p. 69).

[85] Yuan, Quan; Cardei, Ionut; Wu, Jie. Predict and relay: an efficient routing in disruption-tolerant networks. In: *Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing*. 2009, pp. 95–104 (cit. on p. 69).

[86] Daly, Elizabeth M; Haahr, Mads. Social network analysis for routing in disconnected delay-tolerant manets. In: *Proceedings of the 8th ACM international symposium on Mobile ad hoc networking and computing*. 2007, pp. 32–40 (cit. on p. 69).

[87] Li, Qinghua; Zhu, Sencun; Cao, Guohong. Routing in socially selfish delay tolerant networks. In: *2010 Proceedings IEEE Infocom*. IEEE. 2010, pp. 1–9 (cit. on p. 69).

[88] Costa, Paolo; Mascolo, Cecilia; Musolesi, Mirco; Picco, Gian Pietro. Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks. In: *IEEE Journal on selected areas in communications* 26.5 (2008), pp. 748–760 (cit. on p. 69).

[89] Whitbeck, John; Conan, Vania. HYMAD: Hybrid DTN-MANET routing for dense and highly dynamic wireless networks. In: *Computer communications* 33.13 (2010), pp. 1483–1492 (cit. on p. 69).

[90] Song, Libo; Kotz, David F. Evaluating opportunistic routing protocols with large realistic contact traces. In: *Proceedings of the second ACM workshop on Challenged networks*. 2007, pp. 35–42 (cit. on p. 69).

[91] Su, Jing; Goel, Ashvin; De Lara, Eyal. An empirical evaluation of the student-net delay tolerant network. In: *2006 3rd Annual International Conference on Mobile and Ubiquitous Systems-Workshops*. IEEE. 2006, pp. 1–10 (cit. on p. 69).

[92] Jain, Sushant; Fall, Kevin; Patra, Rabin. Routing in a delay tolerant network. In: *Proceedings of the 2004 conference on Applications, technologies, architectures, and protocols for computer communications*. 2004, pp. 145–158 (cit. on p. 70).

[93] Musolesi, Mirco; Hailes, Stephen; Mascolo, Cecilia. Adaptive routing for intermittently connected mobile ad hoc networks. In: *Sixth IEEE International Symposium on a World of wireless mobile and multimedia networks*. IEEE. 2005, pp. 183–189 (cit. on p. 70).

[94] Mtibaa, Abderrahmen; May, Martin; Diot, Christophe; Ammar, Mostafa. Peoplerank: Social opportunistic forwarding. In: *2010 Proceedings IEEE INFOCOM*. IEEE. 2010, pp. 1–5 (cit. on p. 70).

[95] Xie, Xingguang; Zhang, Yong; Dai, Chao; Song, Mei. Social relationship enhanced predicable routing in opportunistic network. In: *2011 Seventh International Conference on Mobile Ad-hoc and Sensor Networks*. IEEE. 2011, pp. 268–275 (cit. on p. 70).

[96] Vu, Long; Do, Quang; Nahrstedt, Klara. 3r: Fine-grained encounter-based routing in delay tolerant networks. In: *2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*. IEEE. 2011, pp. 1–6 (cit. on p. 70).

[97] Komnios, Ioannis; Diamantopoulos, Sotirios; Tsaoussidis, Vassilis. Evaluation of dynamic DTN routing protocols in space environment. In: *2009 International Workshop on Satellite and Space Communications*. IEEE. 2009, pp. 191–195 (cit. on p. 70).

[98] Borrego, Carlos; Garcia, Gerard; Robles, Sergi. Softwarecast: A code-based delivery manycast scheme in heterogeneous and opportunistic ad hoc networks. In: *Ad Hoc Networks* 55 (2017), pp. 72–86 (cit. on p. 70).

[99] Sánchez-Carmona, Adrián; Robles, Sergi; Borrego, Carlos. PrivHab+: A secure geographic routing protocol for DTN. In: *Computer Communications* 78 (2016), pp. 56–73 (cit. on p. 70).

[100] Borrego, Carlos; Sánchez-Carmona, Adrián; Li, Zhiyuan; Robles, Sergi. Explore and wait: A composite routing-delivery scheme for relative profile-casting in opportunistic networks. In: *Computer Networks* 123 (2017), pp. 51–63 (cit. on p. 70).

[101] Deng, Xia; Chang, Le; Tao, Jun; Pan, Jianping; Wang, Jianxin. Social profile-based multicast routing scheme for delay-tolerant networks. In: *2013 IEEE International Conference on Communications (ICC)*. IEEE. 2013, pp. 1857–1861 (cit. on p. 70).

[102] Lee, Uichin; Oh, Soon Young; Lee, Kang-Won; Gerla, Mario. Relaycast: Scalable multicast routing in delay tolerant networks. In: *2008 IEEE International Conference on Network Protocols*. IEEE. 2008, pp. 218–227 (cit. on p. 70).

[103] Hsu, Wei-jen; Dutta, Debojyoti; Helmy, Ahmed. Profile-cast: Behavior-aware mobile networking. In: *2008 IEEE Wireless Communications and Networking Conference*. IEEE. 2008, pp. 3033–3038 (cit. on p. 70).

[104] Hsu, Wei-jen; Dutta, Debojyoti; Helmy, Ahmed. CSI: A paradigm for behavior-oriented profile-cast services in mobile networks. In: *Ad Hoc Networks* 10.8 (2012), pp. 1586–1602 (cit. on p. 70).

[105] Yang, Peng; Chuah, Mooi Choo. Context-aware multicast routing scheme for disruption tolerant networks. In: *Proceedings of the 3rd ACM international workshop on Performance evaluation of wireless ad hoc, sensor and ubiquitous networks*. 2006, pp. 66–73 (cit. on p. 70).

[106] Xi, Yong; Chuah, Mooi Choo. An encounter-based multicast scheme for disruption tolerant networks. In: *Computer Communications* 32.16 (2009), pp. 1742–1756 (cit. on p. 71).

[107] Lee, Sung-Ju; Gerla, Mario; Chiang, Ching-Chuan. On-demand multicast routing protocol. In: *WCNC. 1999 IEEE Wireless Communications and Networking Conference (Cat. No. 99TH8466)*. Vol. 3. IEEE. 1999, pp. 1298–1302 (cit. on p. 71).

[108] Wang, Lan; Olariu, Stephan. A two-zone hybrid routing protocol for mobile ad hoc networks. In: *IEEE transactions on Parallel and distributed systems* 15.12 (2004), pp. 1105–1116 (cit. on p. 71).

[109] Sivakumar, Raghupathy; Sinha, Prasun; Bharghavan, Vaduvur. CEDAR: a core-extraction distributed ad hoc routing algorithm. In: *IEEE Journal on Selected Areas in communications* 17.8 (1999), pp. 1454–1465 (cit. on p. 71).

[110] Pei, Guangyu; Gerla, Mario; Chen, Tsu-Wei. Fisheye state routing: A routing scheme for ad hoc wireless networks. In: *2000 IEEE International Conference on Communications. ICC 2000. Global Convergence Through Communications. Conference Record*. Vol. 1. IEEE. 2000, pp. 70–74 (cit. on p. 71).

[111] Basagni, Stefano; Chlamtac, Imrich; Syrotiuk, Violet R; Woodward, Barry A. A distance routing effect algorithm for mobility (DREAM). In: *Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking*. 1998, pp. 76–84 (cit. on p. 71).

[112] Santivanez, Cesar A; Ramanathan, Ram; Stavrakakis, Ioannis. Making link-state routing scale for ad hoc networks. In: *Proceedings of the 2nd ACM international symposium on Mobile ad hoc networking & computing*. 2001, pp. 22–32 (cit. on p. 71).

[113] Helmy, Ahmed; Garg, Saurabh; Pamu, Priyatham; Nahata, Nitin. Contact-based architecture for resource discovery (CARD) in large scale MANets. In: *Proceedings International Parallel and Distributed Processing Symposium*. IEEE. 2003, 9–pp (cit. on p. 71).

[114] Guangyu, Pei; Geria, M; Hong, Xiaoyan. LANMAR: landmark routing for large scale wireless ad hoc networks with group mobility. In: *2000 First Annual Workshop on Mobile and Ad Hoc Networking and Computing. MobiHOC (Cat. No. 00EX444)*. IEEE. 2000, pp. 11–18 (cit. on p. 71).

[115] Ramasubramanian, Venugopalan; Haas, Zygmunt J; Sirer, Emin Gün. SHARP: A hybrid adaptive routing protocol for mobile ad hoc networks. In: *Proceedings of the 4th ACM international symposium on Mobile ad hoc networking & computing*. 2003, pp. 303–314 (cit. on p. 71).

[116] Heinzelman, Wendi Rabiner; Chandrakasan, Anantha; Balakrishnan, Hari. Energy-efficient communication protocol for wireless microsensor networks. In: *Proceedings of the 33rd annual Hawaii international conference on system sciences*. IEEE. 2000, 10–pp (cit. on p. 71).

[117] Lindsey, Stephanie; Raghavendra, Cauligi S. PEGASIS: Power-efficient gathering in sensor information systems. In: *Proceedings, IEEE aerospace conference*. Vol. 3. IEEE. 2002, pp. 3–3 (cit. on p. 71).

[118] Manjeshwar, Arati; Agrawal, Dharma P. TEEN: ARouting Protocol for Enhanced Efficiency in Wireless Sensor Networks. In: *ipdps*. Vol. 1. 2001. 2001, p. 189 (cit. on p. 72).

[119] Manjeshwar, Arati; Agrawal, Dharma P. APTEEN: A hybrid protocol for efficient routing and comprehensive information retrieval in wireless sensor networks. In: *Parallel and distributed processing symposium, international*. Vol. 3. Citeseer. 2002, 0195b–0195b (cit. on p. 72).

[120] Chakchouk, Nesrine; Hamdaoui, Bechir; Frikha, Mounir. WCDS-DCR: an energy-efficient data-centric routing scheme for wireless sensor networks. In: *Wireless Communications and Mobile Computing* 12.2 (2012), pp. 195–205 (cit. on p. 72).

[121] Biswas, Sanjit; Morris, Robert. Opportunistic routing in multi-hop wireless networks. In: *ACM SIGCOMM Computer Communication Review* 34.1 (2004), pp. 69–74 (cit. on p. 72).

[122] Westphal, Cedric. Opportunistic routing in dynamic ad hoc networks: The OPRAH protocol. In: *2006 IEEE International Conference on Mobile Ad Hoc and Sensor Systems*. IEEE. 2006, pp. 570–573 (cit. on p. 72).

[123] Koutsonikolas, Dimitrios; Wang, Chih-Chun; Hu, Y Charlie. Efficient network-coding-based opportunistic routing through cumulative coded acknowledgments. In: *IEEE/ACM Transactions on Networking* 19.5 (2011), pp. 1368–1381 (cit. on p. 72).

[124] Chachulski, Szymon; Jennings, Michael; Katti, Sachin; Katabi, Dina. Trading structure for randomness in wireless opportunistic routing. In: *ACM SIGCOMM Computer Communication Review* 37.4 (2007), pp. 169–180 (cit. on p. 72).

[125] Rozner, Eric; Seshadri, Jayesh; Mehta, Yogita; Qiu, Lili. SOAR: Simple opportunistic adaptive routing protocol for wireless mesh networks. In: *IEEE transactions on Mobile computing* 8.12 (2009), pp. 1622–1635 (cit. on p. 72).

[126] Yuan, Yuan; Yang, Hao; Wong, Starsky HY; Lu, Songwu; Arbaugh, William. ROMER: Resilient opportunistic mesh routing for wireless mesh networks. In: *IEEE workshop on wireless mesh networks (WiMesh)*. Vol. 12. 2005 (cit. on p. 72).

[127] Zeng, Kai; Lou, Wenjing; Yang, Jie; Brown, Donald R. On throughput efficiency of geographic opportunistic routing in multihop wireless networks. In: *Mobile Networks and Applications* 12.5 (2007), pp. 347–357 (cit. on p. 72).

[128] Zeng, Kai; Lou, Wenjing; Zhai, Hongqiang. Capacity of opportunistic routing in multi-rate and multi-hop wireless networks. In: *IEEE Transactions on Wireless Communications* 7.12 (2008), pp. 5118–5128 (cit. on p. 72).

[129] Lin, Yunfeng; Li, Baochun; Liang, Ben. CodeOR: Opportunistic routing in wireless mesh networks with segmented network coding. In: *2008 IEEE International Conference on Network Protocols*. IEEE. 2008, pp. 13–22 (cit. on p. 72).

[130] Dubois-Ferrière, Henri; Grossglauser, Matthias; Vetterli, Martin. Valuable detours: Least-cost anypath routing. In: *IEEE/ACM Transactions on Networking* 19.2 (2010), pp. 333–346 (cit. on p. 73).

[131] Laufer, Rafael; Velloso, Pedro B; Vieira, Luiz Filipe M; Kleinrock, Leonard. PLASMA: A new routing paradigm for wireless multihop networks. In: *2012 Proceedings IEEE INFOCOM*. IEEE. 2012, pp. 2706–2710 (cit. on p. 73).

[132] Li, Yanhua; Mohaisen, Abedelaziz; Zhang, Zhi-Li. Trading optimality for scalability in large-scale opportunistic routing. In: *IEEE Transactions on Vehicular Technology* 62.5 (2012), pp. 2253–2263 (cit. on p. 73).

[133] Zhao, Jing; Cao, Guohong. VADD: Vehicle-assisted data delivery in vehicular ad hoc networks. In: *IEEE transactions on vehicular technology* 57.3 (2008), pp. 1910–1922 (cit. on p. 73).

[134] Wang, Zehua; Chen, Yuanzhu; Li, Cheng. CORMAN: A novel cooperative opportunistic routing scheme in mobile ad hoc networks. In: *IEEE journal on selected areas in communications* 30.2 (2012), pp. 289–296 (cit. on p. 73).

[135] Yang, Shengbo; Zhong, Feng; Yeo, Chai Kiat; Lee, Bu Sung; Boleng, Jeff. Position based opportunistic routing for robust data delivery in MANETs. In: *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*. IEEE. 2009, pp. 1–6 (cit. on p. 73).

Funding

This PhD studies as well as this PhD thesis were partially funded by Secretaria de Educación Superior, Ciencia, Tecnología e Innovación (SENESCYT, ECUADOR)

Colophon

This thesis was typeset with LaTeX 2_ε. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at `http://cleanthesis.der-ric.de/`.

Declaration

You can put your declaration here, to declare that you have completed your work solely and only with the help of the references you mentioned.

Cerdanyola del Valles, September, 2022

Diego Mauricio Freire Bastidas