ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2015-ENST-0009

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Communication et Electronique »

présentée et soutenue publiquement par

Pavlos SERMPEZIS

le 10 février 2015

Réseaux Sociaux Mobiles

Directeur de thèse : **Thrasyvoulos SPYROPOULOS**

Jury

M. Christian BONNET, Professeur, EURECOM	Président du jury
M. Chadi BARAKAT, Rechercheur, INRIA Sophia-Antipolis	Rapporteur
M. Andrea PASSARELLA, Rechercheur, IIT-CNR Pisa	Rapporteur
M. Jörg OTT, Professeur, Aalto University	Examinateur
M. Ioannis STAVRAKAKIS, Professeur, University of Athens (UoA)	Examinateur

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

T H È S E

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

EURECOM
Sophia Antipolis

DISSERTATION

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
from Telecom ParisTech

Specialization

Communication and Electronics

École doctorale Informatique, Télécommunications et Électronique (Paris)

Presented by

Pavlos SERMPEZIS

Performance Analysis of Mobile Social Networks with Realistic Mobility and Traffic Patterns

Defense scheduled on February 10^{th} 2015

before a committee composed of:

Prof. Christian Bonnet	President of the Jury
Dr. Chadi Barakat	Reporter
Dr. Andrea Passarella	Reporter
Prof. Jörg Ott	Examiner
Prof. Ioannis Stavrakakis	Examiner
Prof. Thrasyvoulos Spyropoulos	Thesis Supervisor

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

EURECOM
Sophia Antipolis

THÈSE

présentée pour obtenir le grade de
Docteur de
Telecom ParisTech

Spécialité

Communication et Electronique

École doctorale Informatique, Télécommunications et électronique (Paris)

Présentée par

Pavlos SERMPEZIS

Réseaux Sociaux Mobiles

Soutenance de thèse prévue le 10 février 2015

devant le jury composé de:

Prof. Christian Bonnet	Président du jury
Dr. Chadi Barakat	Rapporteur
Dr. Andrea Passarella	Rapporteur
Prof. Jörg Ott	Examinateur
Prof. Ioannis Stavrakakis	Examinateur
Prof. Thrasyvoulos Spyropoulos	Directeur de thèse

Abstract

The recent evolution of mobile communications and the wide spread of "smart" mobile devices have radically changed the way we communicate. Ubiquitous access to Internet, fast data rates, online and location-based social networking, are some of the numerous possibilities that can be offered to someone even through his mobile phone. In addition to this, portable devices have become powerful, they support multiple wireless radio interfaces, use many sensors, have large storage capacity, etc. All these capabilities have led to a number of novel services and applications, and new communication paradigms.

Conventional communication between users, through cellular networks or the Internet, can now be complemented by direct communication between mobile devices. Users can directly exchange data with each other using only local wireless communication (e.g. Bluetooth or WiFi Direct), and they can form mobile networks with their peers, in parallel to a cellular or WLAN network, or even when infrastructure is absent. These ad-hoc *Mobile Social Networks* (MSNs) can support communication in challenging environments, where infrastructure is limited (e.g. emergency situations after disasters, rural areas), or enhance existing networking infrastructure, e.g. by offloading traffic from cellular networks, enabling novel social and location-based applications, or introducing peer-to-peer collaborative computing.

In MSNs, a message can be directly delivered to the destinations when they meet with the source node(s) (single-hop) or relay-assisted schemes are employed, where relay nodes *store* the message, *carry* it as they move and (possibly) *forward* it to other relays till it reaches its destination (multi-hop). Since mobile-to-mobile communication takes place only during meetings (*contacts*) between nodes, the communication performance in MSNs heavily depends on the underlying node mobility patterns. Communication traffic patterns (i.e. who *wants* to communicate with whom or *is interested* in what) can significantly affect the performance of communication mechanisms too. In addition, numerous studies from different disciplines, like sociology, opportunistic networking, social media etc., have shown that social characteristics of users affect both their mobility and traffic patterns. As a result, nodes' different social behaviors can lead to very heterogeneous MSNs.

To this end, the primary focus of the thesis is on understanding, *analytically*, to what extent social heterogeneity affects the performance of the different networking solutions (e.g. forwarding/routing protocols or content-centric schemes) in MSNs. Towards this direction, we propose models that take into account key aspects of realistic mobility and traffic patterns, but, simultaneously, remain simple enough to allow tractable analysis. The second goal is to propose, based on this analysis, some general design guidelines and/or insights about communication protocols.

Specifically, after providing in Chapter 1 a short introduction to MSNs and the motivation of our work, we first study the effects of heterogeneous mobility patterns in Chapter 2. We define a class of models for *Heterogeneous Contact Networks*, i.e. networks where nodes' mobility

(and, thus, also their meeting/contacting patterns) is heterogeneous. These models capture two basic characteristics of real mobility, namely (i) different node pairs can contact with different frequency, and (ii) some node pairs never contact each other. We perform an asymptotic analysis for the delivery delay of a message spreading, and derive closed form results that can be used to predict the delay of epidemic-based routing protocols.

Although contact events between nodes denote the times when a message can be exchanged between two nodes, one should also take into account that not all nodes are always willing to relay third party traffic (as most protocols assume) in a MSN. Hence, some contact events might be incorrectly considered as opportunities for message exchanges. In addition, such reluctance from nodes to cooperate (node selfishness), is usually related to the social ties between them, e.g. it might be more probable to relay a message generated by a friend node, rather than from an unknown device. To this end, in Chapter 3, we propose a model that captures *social selfishness* through its correlation to mobility patterns. We extend the analysis of Chapter 2 and derive results that show how social selfishness impacts the efficiency of various communication mechanisms.

After having studied the effects of heterogeneity in contact events between (cooperative) nodes, in Chapters 4 and 5 we turn our attention to communication traffic patterns, whose effects have not been previously studied analytically in MSNs. In particular, in Chapter 4 we investigate when traffic heterogeneity can affect performance, we propose a generic model to describe generic traffic patterns, and we derive analytic results that quantify the joint effects of traffic and mobility heterogeneity on *end-to-end communication* mechanisms.

On a different direction, in Chapter 5 we consider *content-centric communications*, where a certain message (content) needs to be distributed to nodes (more than one) that are interested in it. We model and study the two main factors of traffic (or interest) patterns that affect communication, namely the content popularity (how many nodes are interested in a content) and availability (how many nodes can provide a content). We derive useful expressions for the average content delivery delay and the delivery probability by a given deadline, and use them to optimize the performance of a mobile data offloading scheme.

Based on the insights stemming from our analysis in Chapter 5, we focus on mobile data offloading in Heterogeneous Networks (HetNets) comprising infrastructure (small-cells) and mobile (MSN) edge nodes. In Chapter 6, we model and analyze the content dissemination, and calculate the performance of the system, as well as the costs it incurs for the cellular network operator. We then formalize the offloading cost minimization problem and provide initial insights for optimal storage allocation policies.

Finally, we conclude our findings and discuss future research direction in Chapter 7.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Thrasyvoulos Spyropoulos. This work would not have been possible without his continuous guidance and support. He is a great teacher that patiently guided me through research principles during these three last years. Working with him was an enjoyable experience and an invaluable lesson for my future professional life.

I would also like to thank the people from our research group, Fidan, Nikos, Delia, Luigi, with whom I had the pleasure to collaborate. They were always willing to discuss about research problems and offer me their advice.

Leaving my country, moving to a new place and entering a new, often stressful and demanding, working environment, seemed to be difficult. However, being surrounded by friends, made things much easier, and gave me the power to overcome any difficulties. Thus, many thanks go to my friends that I met here and shared with them unforgettable moments: Miltos, Iraklis, Panos, Giorgos, Hadrien, Akis, Valia, Nikos, Giorgos, as well as to Dora, Petros, Maria, Froso, Miriam, Jonas, Martina, Konstantinos; and also to my friends who were far, but with whom we had a real communication, through our short skype discussions or during the few days of the year that we could meet: Vasilis, Paschalis, Nikos, Marina, Danai, Maria, Babis, Giannis, Lefteris, Nikos, Mitsos.

I am extremely grateful, and I dedicate this thesis, to my parents, Dimitris and Maria, and my sister Ioanna for always loving and supporting me, and for everything they have offered me in my life.

Finally and *most importantly*, my special thanks goes to Niki for standing next to me, despite the difficulties, and for her constant love, understanding, and patience all these years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Mobile Social Networks (MSNs)

The recent advances in mobile communication technologies have led to faster connection speeds, ubiquitous connectivity and access to information resources, countless mobile applications, transforming thus continuously the way we communicate, and even affecting many habits of our everyday life. The prevalence of online social networking, or the integration of a multitude of services (entertainment, navigation, etc.) in a single handheld device, are prominent examples of this ample and ongoing evolution.

The proliferation of portable devices with augmented capabilities has largely contributed towards these changes. Modern portable communication devices, like smartphones, pads, laptops, are equipped with a number of communication interfaces (3G/4G, WiFi, Bluetooth, etc.), large storage capacity, and high computational power. As a result, communication is not restricted to traditional voice calls or messaging, but it is enriched with new elements, e.g. rich multimedia sources, introduced by novel applications. Moreover, the increased density of mobile devices and the ability to mingle different wireless communication techniques, enables new ways of mobile networking. Combining cellular communications (through base stations), short-range communications with the infrastructure (e.g., WiFi access points) or directly between neighboring devices, in an ad-hoc manner (e.g. using Bluetooth, WiFi Direct), a user can enhance its connectivity to Internet, connect and exchange data with its peers, share data (content sharing) or resources (collaborative computing, mobile cloud computing), etc.

Through this broad potential, the *Mobile Social Networking* paradigm has emerged. In Mobile Social Networks (MSNs), the term "Mobile" indicates that end nodes are users connected to the network through a mobile device (infrastructure nodes, e.g. acting as relays or gateways, can be a part of the network as well), while the term "Social" indicates their focus, which is mainly on social networking applications, and/or the fact that the participants' social characteristics are exploited in order to set up, facilitate, or enhance communication.

Although the social dimension has been recently introduced in mobile networking with MSNs, the mobility component was inherited from previous networking paradigms. Specifically, and following a chronological order, one can think of Mobile Ad-hoc Networks (MANETs), Delay Tolerant Networks (DTNs), and Opportunistic Networks, as the ancestors of MSNs. MANETs are ad-hoc networks, where nodes can move, causing thus a frequently changing topology. When topology has changed, nodes have to re-calculate (in a distributed or local manner) the connec-

tivity graph, the multi-hop routing paths, etc. Delay Tolerant Networking (DTN) has been proposed a decade ago as an architecture for challenged networks [33]. Despite the resemblance to MANETs, DTNs are regarded as an individual research field due to a number of different major characteristics: disconnections in communication between nodes, frequent absence of (continuous) end-to-end paths, long delays and low throughput, etc. [33]. The term "Opportunistic Networks", although it has been introduced later than the term "DTNs", is often used interchangeably with DTNs, or, according to [110], it describes a more generic class of intermittently connected networks.

From the *network connectivity perspective*, MSNs can be considered equivalent to Opportunistic Networks or DTNs. A difference is that while DTNs might refer to sensor networks, vehicular networks, deep space communication networks, etc., and Opportunistic Networking has been used to describe a wide range of networking environments [110], from pocket switched networks [58] to wildlife monitoring [67, 128], Mobile Social Networks are mostly used for describing networks composed of portable devices (as well as infrastructure nodes) or refer to applications used by mobile phone users. The common baseline among these networks (DTNs, Opportunistics Networks, and MSNs) is the way their nodes can connect and communicate with each other. MSNs (or DTNs, Opportunistics Networks) are composed mainly of nodes moving in an area much larger than their transmission range. Data exchange between nodes can take place only when they are within transmission range of each other, or, as it is also called, when they are *in contact*. Message dissemination can be end-to-end or content-centric, yet neither the existence nor the knowledge of an end-to-end path is assumed. Message dissemination from a source to a destination node could be achieved by direct transmission [130], when source and destination come in contact. Alternatively, over a sequence of node encounters, messages can get copied to many nodes, stored and carried by them (as nodes move), and forwarded over multiple hops to the destination.

Early opportunistic networking solutions comprise flooding mechanisms, where a message is copied by every node having it to every node not having it upon their encounter, and epidemic-based limited replication schemes, where a message can be copied to a maximum (concurrent) number of intermediate nodes (we refer to them as *relays*) before it reaches its destination(s). With the advent of MSNs, the design of routing methods has advanced: protocols exploit the knowledge of nodes' *social characteristics* in order to make better forwarding decisions and achieve faster (and/or more likely) message delivery. Furthermore, the social component had an effect on the envisioned use cases for MSNs: a number of novel applications have been proposed, which merge mobile networking with social networking, and fit better to network environments composed mainly of people holding portable devices (rather than other kind of non-rational nodes, e.g. sensors).

In the remainder, we give a general overview of the aforementioned use cases and networking solutions for MSNs, as a preliminary step towards understanding (a) the research challenges in MSNs that motivated our work, and (b) what are the contributions of this thesis.

1.1.1 Use Cases

Extreme Environments. The initial application of Opportunistic Networks / DTNs was to support communication in challenging environments with total or partial absence of infrastructure. Some prominent examples, which apply in the more specific case of MSNs as well, are:
– Providing asynchronous connectivity to Internet or between users of the same network, in

rural or developing areas, where conventional networks are not deployed (and it is not feasible or cost efficient to do so), see e.g. [42,46,111]. Special stations can be installed in remote villages, and users request information from these stations. If the requested content does not exist in the station's storage, then the station requests it from mobile access points (e.g. mounted on vehicles) or other users (e.g. with smartphones), which periodically pass by the stations and other locations where connectivity to Internet is available.
– Allowing users to communicate in emergency from spots without infrastructure. For instance, in situations after disasters that destroyed infrastructure [53], or areas without mobile coverage [136], people that need help (injured, trapped, etc.) can wirelessly communicate with neighboring devices, which will then spread the data from phone to phone till it access a rescue team or a region where infrastructure is operational.

Mobile Data Offloading. The mobile data demand is rapidly increasing, due to the recent growth in the number of mobile devices and connection speeds. Cellular networks are currently overloaded and they are not expected to be able to keep up with the data demand [23]. As a result, mechanisms that reduce cellular traffic by offloading data to mobile devices have attracted a lot of attention. In opportunistic mobile data offloading, e.g. [50,85,141], the cellular network provider, instead of serving separately each user requesting a content (e.g. a popular video, or software update), distributes a few copies of the content in some relay nodes, which store it in their caches. A user interested in the given content, can retrieve it through direct communication from a relay node holding it, when they come within transmission with each other. Although this mechanism implies that the users might have to wait for some amount of time until they receive the content (i.e. they encounter a node storing it), appropriate incentives (e.g. price reductions) [48] can be provided by the operator to guarantee the feasibility of the service.

Location-based Applications. In many mobile applications, the specific data that is accessed depends on or is related to the current location of the user [62, 105, 133]. Examples include live road traffic information, reviews of restaurants and local businesses, local event notification, map tiles, localization services, etc. To access such data, a user has first to obtain its location (e.g. through GPS) and transmit it to the data provider. However, this might rise location or content privacy concerns, and overload wireless access links unnecessarily. Alternatively, users that reside in the same area and frequently come within transmission range can form a MSN, and exchange directly and spread such location-based information.

Mobile Computing. Every node in a MSN is a powerful mobile device, with large memory and processing power, and a number of sensors. Combining the software, hardware, and sensing resources of more than one devices allows to increase their capabilities, by building a mobile cloud, distributed applications, or collaborative sensing [26,119,122,123]. For instance, tasks that cannot be executed in a single device of a MSN, because not all resources (e.g. sensors, software) are available in it, or they need a lot of time, or they are energy-consuming, can be segmented in sub-tasks. Then the device can assign the sub-tasks to other devices, by sending them the necessary input through the MSN (single-hop direct transmission or multi-hop communication using relay nodes). When a sub-task is executed, the output is send back to the requesting device through the MSN as well. After the completion and reception of all sub-task outputs, the results are combined to complete the service.

1.1.2 Networking Solutions

The above use cases show that MSNs can support both *end-to-end* and *content-centric* applications. In end-to-end applications, a source node generates a message (or a sequence of messages), whose destination is a specific node (unicast) or a set of specific nodes (multicast). Examples of end-to-end communication can be online social networking (e.g. exchange of Facebook messages, or posts on Twitter), and emergency messages sent from a user needing help to all rescue stations/nodes. On the other hand, in the content-centric communication model, many users are interested in the same content or, possibly, in any content belonging to a given category. A "content" can be a message (i.e. a data file, like a map tile, a trending video, etc.) or even a service provided by other users (see Mobile Computing applications). Interested users can access the content they are looking for directly from any encountered node that offers it.

End-to-end communication. A message exchange between two users in a MSN can be achieved (i) by direct transmission, where the source transmits the message directly to the destination when they come within transmission range, or (ii) through the *store-carry-forward* mechanism, where nodes of the network can be used as relays and, after they receive the message from the source (or another relay), they can deliver it to the destination, when they contact, or forward it to other relays.

Since message exchanges can take place only during contact events between nodes, it becomes evident that the communication performance (e.g. how fast a message can be delivered, or what is the probability of never reaching its destination) heavily depends on the mobility of the nodes involved in the communication process (i.e. the source, destination, and relay nodes). When information about the mobility patterns is not available, the selection of the relay nodes is random [43, 129, 137, 143], and one can just use the maximum number of copies (i.e. concurrent relays) as a protocol design parameter [129]. Nevertheless, in an MSN, some social characteristics of its participants are known, and thus information about their mobility patterns can be usually retrieved as well: either explicitly, e.g. in cases where users share mobility related data with other users or with a central entity [30], or implicitly, e.g. when users belong to a social network and their mobility is inferred by other correlated available data, like their social ties [54]. Therefore, recent MSN protocols exploit such mobility related information to improve delivery performance[1], for instance, by selecting relay nodes that are expected to meet soon the destination.

Socially-aware routing protocols. Although some information about mobility patterns is available, neither the exact movement trajectories of nodes nor the contact events between them can be known a priori. Hence, a deterministic calculation of the optimal paths is not possible in MSN routing. Mobility-related information that is usually available or a user is willing to disclose is (i) how *frequently* contacts with other users, (ii) the *duration* of staying in contact (i.e. within transmission range of each other), or (iii) the time of *last encounter* with every of them. The goal of a social-aware MSN routing algorithm is to use such history-based knowledge of mobility metrics, and predict future contact events between nodes. Then, based on these predictions, the optimal set of relays or the forwarding and routing policies can be selected in order to improve performance.

[1] However, we need to stress here that the base mechanism behind all these protocols derive from the store-carry-forward framework and epidemic-based spreading.

A common way to capture such mobility and interactions between nodes is through graph representation, namely the *contact graph*, which is defined as following

Definition 1 (Contact Graph)**.** *The contact graph of a network $\mathcal{N}$ is a weighted graph $\mathcal{G} = \{V, E\}$ whose vertices represent the network nodes and an edge between two vertices implies that these two nodes can contact each other regularly; the weight of an edge measures the mobility correlation between the respective nodes.*

Different techniques have been proposed for how to build a contact graph from the knowledge of past contact events, e.g. which metrics should be used (frequency, duration, age of contacts) and how to infer the edge weights from them (e.g. aggregation or sliding time window) [35, 52].

Finally, after having built the contact graph, a node can select to which node to forward a message depending on the weight of the edge connecting it with the destination (e.g. EBR [98]), or their similarity (which relates to community structure), or its centrality (see e.g. SimBet [29] or BubbleRap [60]), etc.

Content-centric communication. The underlying mechanism for disseminating data in content-centric applications is the same as in end-to-end applications: content is delivered to interested users through direct transmissions or store-carry-forward schemes. Therefore, techniques that exploit nodes mobility patterns are also used in content-centric protocols. For instance, the holders of a content (i.e. the nodes which are delegated to distribute the content) can be selected according to the weights of edges connecting them with every node interested in the content [34, 85], or based on the community structure of the network [10, 142].

In addition to mobility, content-centric communication is inherently related to the interests of users. Thus, to optimally select the content holders, it does not suffice to know how nodes move, but information about their interests is needed as well. Usually users interests are not explicitly known, and protocols are based on predictions of interests, e.g. inferring interests patterns from the social relationships between users, their association to social communities, etc., [10, 28, 142].

1.2 Motivation and Contributions of the Thesis

1.2.1 Motivation

Performance Evaluation. The goal of a MSN communication mechanism is to efficiently deliver data to the destination nodes (end-to-end) or to any interested node (content-centric). To quantify this efficiency, metrics like the *delivery delay* (how fast nodes receive a message or find a content of interest), *delivery probability* (how probable for a message/content is to reach a node by a given deadline), and *overhead* (e.g. number of relays or transmissions) per message, are used. The performance evaluation of MSN protocols, as well as the comparison of different approaches, is usually done by calculating the statistics of such metrics through simulations, real experiments, or analytic models.

In simulations, a trace of node movements is generated, and the timings of communication opportunities between them (contact events) are calculated. Mobility traces are usually generated by random (uniform) mobility models, like the Random Walk or Random Waypoint models, or, more recently, by state-of-the-art mobility models, e.g. [13, 14, 57, 77, 93, 97], that capture the complex, heterogeneous characteristics of mobility patterns observed in real traces [25,35,36,117].

Although simulations can provide with accurate results in certain settings, they can be lengthy and complex when the performance needs to be evaluated under a various range of network parameters. Scalability problems appear in performance evaluation through real experiments as well. Furthermore, till now large-scale MSNs have not been deployed and the only experience we have is of small, experimental settings [19, 31, 51, 58, 81, 91], which might not always satisfy the necessary conditions for a thorough performance evaluation of different protocols. Hence, the need for performance prediction under generic settings, led researchers to study MSNs using analytic models. Analytic models not only can be used for performance prediction, but also can provide useful insights about the feasibility of possible applications, appropriate selection of routing mechanisms, etc., and reveal which are the network characteristics/parameters that affect performance and to what extent.

Analytic Models. Early analytic models were based on simple mobility assumptions [43, 49, 143], namely:

- The sequence of contact events between a pair nodes is given by a *Poisson process*; or equivalently, the inter-contact times (i.e. time intervals between two successive contacts) are independent and exponentially distributed.
- Mobility is homogeneous, with *every node pair* contacting with the *same frequency* (i.e. with rate λ).

With the above assumptions, the message dissemination can be modeled using absorbing *Markov Chains* [43,49] or *fluid models* [49,143]. This simplifies analysis and closed form expressions predicting the message delivery delay, delivery probability, overhead per message, buffer occupancy, etc., can be found for a number of epidemic-based schemes. Simple, closed form expressions not only facilitate performance evaluation, but also provide useful intuition for the dependence between communication performance and network characteristics. For example, with only a single inspection of the expressions in [43, 143], one can see how performance metrics change with network parameters, like the total number of nodes or the contact rate λ.

However, the social characteristics of users (including their mobility patterns) in a MSN cannot be expected to be homogeneous [25, 35, 36, 117], rendering thus the above assumptions unrealistic. Moreover, the large number of social-aware protocols (see Section 1.1.2), which exploit the network heterogeneous characteristics, cannot be analyzed with the above homogeneous models where all nodes (or node pairs) are considered equivalent. Motivated by this insufficiency of previous models, a number of more realistic analytic models capturing observed social properties and mobility patterns have been proposed [9,11,20,34,36,63,73,76,113,114,132].

Some main modeling approaches followed in these studies are to consider networks with heterogeneous contact rates, where each node pair $\{i, j\}$ contacts with rate λ_{ij} (which can be different among different pairs), e.g. [36, 113, 114], or to divide nodes in social or spatial communities, where nodes residing in the same community contact each other with the same rate, but contact rates in different communities (or between two communities) can take different values [20, 73, 76, 132]. However, introducing heterogeneous rates, increases the complexity of analyzing the performance of content dissemination schemes. As a result, exact predictions of performance metrics cannot be derived in closed form expressions, and only numerical solutions, e.g. [73, 132], or upper bounds using rough spectral arguments [113] are allowed. Hence, despite the usefulness of these models in predicting performance under a given setting and/or a given

dissemination scheme, it is not always possible to generalize their findings or to provide insights about how communication performance is affected by the different network characteristics.

To this end, in this thesis, we propose analytic models that take into account key aspects of heterogeneity in MSNs, and, simultaneously, remain simple enough to allow tractable analysis and derivation of closed form results. Our aim is to (i) provide useful intuition about how performance is affected by the different network parameters (e.g. network size, traffic patterns, cooperation of nodes) when mobility is heterogeneous, as well as how these effects change under varying mobility heterogeneity, and (ii) propose some general design guidelines about routing protocols and content-dissemination mechanisms.

In the next section, we summarize the contributions and present the outline of the thesis.

1.2.2 Contributions and Outline

The focus of this thesis is on understanding, *analytically*, the effects of social heterogeneity on the performance of information dissemination mechanisms. The different social characteristics of people (i.e. the users in a MSN), affect a number of aspects related to the communication performance in MSNs, like the way they move (which places visit more frequently, with whom they contact regularly, etc.) and communicate (with whom and how frequently), their interests, their willingness to participate in data dissemination, etc. Throughout the remaining chapters, we address such issues, by (i) trying to capture different dimensions of heterogeneity with analytic models, (ii) analyzing the effects on the performance of basic communication mechanisms, (iii) providing useful intuition and discussing important implications for mobile social networking. Since models and analysis are usually based on simplifying assumptions, which might not be always able to capture exactly all complex characteristics of a real MSN, we test the validity of our results through extensive simulations in a number of network settings.

Specifically, the chapters of the thesis, and the main contributions in each of them, are organized as following:

Chapter 2 – Delay Analysis of Epidemic Schemes in Sparse and Dense Heterogeneous Contact Networks.

As stressed earlier, nodes can exchange data only when they are in contact. Therefore, to analyze communication, we first need a model describing the way nodes contact each other. To this end, in Chapter 2, we define a class of *heterogeneous contact models* (or, mobility models), with which we can capture heterogeneity in contact processes among different nodes and node pairs. In particular, we extend previous homogeneous models (see Section 1.2.1) in the following two directions: (i) we allow different node pairs to contact each other with different frequencies; (ii) we allow some pairs to never contact each other.

With respect to realism and simplicity, the class of models we define, lies between homogeneous models (unrealistic, but simple), e.g. [43, 49, 143], and previously proposed heterogeneous models (realistic, but complex), e.g. [36,73,113]. As a result, analysis becomes tractable (though complex) even for settings with heterogeneous contact/mobility patterns, and this allows us to derive simple, closed form expressions that require knowledge of only a few network parameters, for the information spreading delay in a network. To demonstrate the utility of these results in practice, we use them to compute the delay of basic epidemic-based schemes.

The works related to this chapter are:

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Delay analysis of epidemic schemes in sparse and dense heterogeneous contact environments", Research Report RR-12-272, Eurecom, July 2012.*

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Information diffusion in heterogeneous networks: The configuration model approach", Proc. 5th IEEE International Workshop on Network Science for Communication Networks (NetSciCom'13), co-located with IEEE INFOCOM 2013, 19 April 2013, Turin, Italy.*

Chapter 3 – Understanding the Effects of Social Selfishness.

Mobility models determine when nodes are in contact, and assuming that data exchanges take place during these contacts, one can evaluate various routing and forwarding algorithms. However, any possible unwillingness of the relay nodes to cooperate (i.e. to store or forward a message at a contact event) can affect gravely the performance of message dissemination techniques. Hence, it becomes evident that performance is controlled only by a subset of the contact events; those in which nodes are willing to exchange data.

To this end, in this chapter, we study analytically the effects of node cooperation, or node *selfishness*, on mobile social networking. Extending previous studies that assumed uniform selfishness patterns, e.g. nodes are equally reluctant to cooperate, we propose a framework for analysing cases where the level of selfishness, is related to social ties between nodes or their mobility patterns. We refer to this, correlated to social characteristics, selfishness as *social selfishness*. Incorporating our social selfishness model into the models of Chapter 2, we capture the combined effects of mobility and selfishness heterogeneity. Following also a similar analysis, we derive expressions for important metrics, namely the message delivery delay, the average power consumption and the message delivery probability, and demonstrate the applicability of our results in various application scenarios.

The work in this chapter is published in:

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Understanding the effects of social selfishness on the performance of heterogeneous opportunistic networks", Computer Communications, Elsevier, Volume 48, April 2014.*

Chapter 4 – Modeling and Analysis of Communication Traffic Heterogeneity in MSNs.

Despite the fact that mobility heterogeneity and its impact in MSNs has been extensively studied (through simulations or analyses), this has not been the case with *communication traffic patterns*. In the vast majority of works on performance evaluation of MSN routing protocols, traffic is assumed to be homogeneous, i.e. each pair of nodes is equally probable to be the source and destination of a message. This assumption is generally not true, as nodes' social characteristics can significantly affect the end-to-end traffic demand between them.

Motivated by this lack of related work, in this chapter we explore the effect of *heterogeneous traffic patterns* in MSNs. Based on previous knowledge on the relation between traffic patterns and social characteristics of users, we propose a model to describe traffic heterogeneity. We derive results showing the joint effects of traffic and mobility patterns on end-to-end communication mechanisms. Among the different insights stemming from our analysis, we identify

conditions under which heterogeneity renders the added value of using extra relays more/less useful. Furthermore, we confirm the intuition that an increasing amount of heterogeneity closes the performance gap between different forwarding policies, making end-to-end routing more challenging in some cases, or less necessary in others. We believe these first analytical results on the effects of traffic heterogeneity are an important step towards better protocol design and evaluation of the feasibility of applications in opportunistic networks.

The work in this chapter has resulted to the following submission:

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Modelling and analysis of communication traffic heterogeneity in opportunistic networks", IEEE Transactions on Mobile Computing, pending major revision, October 2014.*

Chapter 5 – Content-Centric Traffic: Effect of Content Popularity and Availability Patterns

In Chapter 4 we focused on the effects of traffic heterogeneity in end-to-end communication. Yet, as discussed in Section 1.1.2, many MSN applications are content-centric: traffic is not between a source-destination node pair, but the main goal is to distribute contents to interested users. Thus, in this case, heterogeneity cannot be viewed from a node pair perspective, but it appears among the groups of nodes involved in the dissemination of different contents. Specifically, the interest patterns, i.e. how many nodes are interested in each content (*popularity*), as well as how many users can provide a content (*availability*), impact the performance and feasibility of content-centric applications.

To this end, in this chapter, we establish an analytical framework to study the effects of these factors on the delay and success probability of a content access request served through mobile social networking. We derive results that calculate these effects as a joint function of (i) mobility patterns, (ii) content popularity patterns, and (iii) content availability patterns. We also derive closed form expressions for performance prediction that require little knowledge of the network characteristics and interest patterns, and thus can be used in real settings, for protocol tuning, online optimization, etc. As an example case, we further apply our framework to the mobile data offloading problem and provide some initial insights for the optimization of its performance.

The work in this chapter corresponds to the following conference paper and its extended version (under submission):

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Not all content is created equal: Effect of popularity and availability for content-centric opportunistic networking", Proc. 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'14), August 11-14, 2014, Philadelphia, PA, USA.*

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Effects of content popularity in the performance of content-centric opportunistic networking: An analytical approach and applications", IEEE/ACM Transactions on Networking, submitted, September 2014.*

Chapter 6 – Offloading on the Edge: Analysis and Optimization of Local Data Storage and Offloading in HetNets

Based on the analysis in Chapter 5 for the effects of content-centric traffic heterogeneity, and the insights stemming from the corresponding results, in this chapter, we focus on a content-centric application, namely *mobile data offloading*, which has recently attracted a lot of attention, due to the rapid increase in data traffic demand that has overloaded cellular networks.

We propose an analytical model to explore how much local storage and opportunistic communication through "edge" nodes could help offload traffic in various heterogeneous network (HetNet) setups and levels of user tolerance to delays. We derive results predicting the performance from the perspective of the user (content delivery probability, delivery delay) and the cellular operator (offloading cost). We then use our model and results to optimize the storage allocation and access mode of different contents as a tradeoff between user satisfaction and cost to the operator.

The work related to this chapter is found in:

- *Pavlos Sermpezis, Luigi Vigneri, Thrasyvoulos Spyropoulos, "Offloading on the Edge: Analysis and optimization of local data storage and offloading in HetNets", Research Report RR-14-297, Eurecom, December 2014.*

Chapter 2

Delay Analysis of Epidemic Schemes in Sparse and Dense Heterogeneous Contact Networks

2.1 Introduction

Epidemic spreading is probably one of the most popular bio-inspired principles that have made their way into computer engineering. Epidemic algorithms and variants have been used for communication in distributed systems, synchronization of distributed databases, content searching in peer-to-peer systems, etc. Recently, epidemic-based schemes have also been proposed for routing and data dissemination in Mobile Social Networks.

In the *epidemic routing* case [137], any node that has a message (is "infected") will forward it to any node encountered that does not have it yet (is "susceptible"). While this guarantees that every node in the network will eventually receive the message, it comes with a high resource overhead. Numerous variants have been proposed to improve the resource usage of epidemic routing while maintaining good performance (see [110, 131] for a detailed survey).

Since the mobility process of nodes involved (e.g. humans or vehicles carrying the devices) is, in most cases, not deterministic, the performance of epidemic-based algorithms heavily depends on the underlying contact patterns between nodes. To this end, epidemic algorithms have been extensively studied through both simulations and analytical models. While simulations with state-of-the-art synthetic models or real mobility traces can provide more reliable predictions for the *specific* scenario tested, analytical models can give quick, qualitative results and intuition, answer "what-if" questions, and help optimize epidemic-based protocols (e.g. choosing the number of copies in [129], or gossip probability [143]).

For the sake of tractability, analytical models for epidemic spreading mainly rely on simple mobility assumptions (e.g. Random Walk, Random Waypoint), where node mobility is stochastic and independent, identically distributed (IID) (see e.g. [43, 49, 143]). Nevertheless, numerous studies of real mobility traces [25, 36, 56, 108] reveal a different picture: Two key findings are that (i) contact rates between different pairs of nodes can vary widely, and (ii) many pairs of nodes may never meet. This puts in question the accuracy and utility of these homogeneous models' predictions. Yet, departures from these assumptions [12, 36, 73, 76, 132] seem to quickly increase complexity and/or limit the applicability of results.

These observations leave us with the following question: *Can we still derive useful and accurate closed-form expressions for the performance of epidemic schemes, even when considering more generic mobility assumptions?*

To this end, in this chapter, we consider a large class of contact/mobility models with heterogeneous contact rates and contact graphs. At first, we assume that every pair of nodes $\{i, j\}$ meets according to a random process with a different contact rate λ_{ij}, drawn from an arbitrary probability distribution $f_\lambda(\lambda)$ with known mean μ_λ and variance σ_λ^2 (Section 2.2):

- Through an asymptotic analysis for the epidemic spreading process, we derive results for the expected spreading delay and provide intuition about how it is affected by the contact rates heterogeneity (Section 2.2.3).
- For finite network sizes, we derive approximative results that predict the expected epidemic spreading delay (Section 2.2.4). The expressions we provide are simple, closed form and only involve the 1^{st} and 2^{nd} moments of the contact rate distribution $f_\lambda(\lambda)$.
- To demonstrate how our framework could be used in practice, we derive closed form expressions for the delay of various epidemic based protocols (Section 2.2.5):

We then further extend the class of mobility models and consider arbitrarily sparse networks by allowing pairs of nodes to never meet each other (Section 2.3):

- Extending the heterogeneous contact model of Section 2.2, we show how the delay predictions we derived can be used for arbitrarily sparse networks modeled as Poisson Graphs (Section 2.3.1).
- We capture further complex characteristics of the network contact graphs, namely heterogeneous node degree distributions, using the Configuration Model. We show (though under uniform contact rates) that we can still derive simple, closed form approximations for various quantities related to the delay of epidemic spreading (Section 2.3.2).

We validate all our results against various synthetic simulation scenarios, and show that their accuracy is significant. Moreover, we test our theory against real traces, capturing node mobility and respective contacts, and find that useful levels of accuracy can still be achieved even for scenarios that are known to entail considerable more complexity.

As a final remark, while our initial motivation and focus stems from the area of MSNs, we believe that our methodology and results could also be applicable to other processes and complex networks [102], if the key metric of interest is spreading delay. In such contexts, contacts between nodes might still be subject to a random process, e.g. related to online communication, email transmission, etc., superimposed over a complex network (e.g. an Online Social Network friendship graph).

2.2 Heterogeneous Contact Networks and Epidemic Schemes

2.2.1 Modeling Heterogeneous Contact Networks

We consider a network $\mathcal{N}$, with N nodes. We assume that the node transmission range is much smaller than the total network area, so that each pair of nodes can only communicate directly during the *contact events* of this pair (i.e. when the two nodes come into the transmission range

of each other). We model this sequence of contact events for a pair of nodes $\{i, j\}$ by a *random point process*[1], and we introduce the concept of the *Contact Network*.

Definition 2 (Contact Network)**.**
– A contact network $\mathcal{N}$ is defined by an (underlying) graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ whose vertices represent the network nodes and an edge between two vertices implies that these two nodes contact each other regularly.
– The sequence of the contact events between each pair of nodes $\{i, j\}$ connected with an edge ($\{i, j\} \in \mathcal{E}$), is given by a random point process with rate λ_{ij}.
– Contact duration is negligible compared to the time between contacts events, though sufficient for all data transfers to take place.

Remark: From the above definition, it follows that, equivalently to its graph, a contact network $\mathcal{N}$ can be represented by the contact matrix $\mathbf{\Lambda} = \{\lambda_{ij}\}$.

Although Def. 2 is quite general, to ensure analytical tractability it is commonly assumed, either implicitly or explicitly, that (i) the underlying graph $\mathcal{G}$ is fully meshed (i.e. $\lambda_{ij} > 0,\ \forall\{i, j\}$), and (ii) mobility is homogeneous (i.e. $\lambda_{ij} = \lambda,\ \forall\{i, j\}$) [43,143]. Yet, in most scenarios of interest, this homogeneity is rather unrealistic. Study of real traces has provided strong evidence that contacts between different pairs of nodes are in fact largely *heterogeneous*, with some pairs never meeting each other and others meeting much more frequently [25,36,108], resulting in a *sparse*, and largely *heterogeneous* contact graph $\mathcal{G}$. This motivates us to depart from the homogeneous mobility model.

Hence, our goal is to extend previous *analytical* works, by considering *heterogeneous contact networks*, and derive useful, closed form results also for these more realistic settings. To this end, we first raise the mobility *homogeneity* assumption. Later, in Section 2.3, we extend our analysis and results for sparse networks (i.e. $\lambda_{ij} \geq 0$) as well. Specifically, we perform our analysis assuming the following *class* of heterogeneous contact networks:

Definition 3 (Heterogeneous Contact Network)**.**
A heterogeneous contact network is defined as a Contact Network (Def. 2), where:
– Contact events between a pair of nodes $\{i, j\}$ follow a Poisson process with rate λ_{ij}, i.e. inter-contact times are independent and exponentially distributed with rate λ_{ij}.
– Contact rates λ_{ij} are independently drawn from an arbitrary distribution with probability density function $f_\lambda(\lambda), \lambda \in [\lambda_{min}, \lambda_{max}] \subseteq (0, \infty)$, and finite mean μ_λ and variance σ_λ^2 (coefficient of variation $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$).

With the choice of the above model we try to strike a tradeoff between realism and usability. We will now motivate our choices above in a bit more detail.

First, the assumption of independent and exponentially distributed inter-contact times (or equivalently Poisson contact processes) for each pair of nodes is needed to allow an exact analysis of performance metrics of interest using a Markovian framework. For this reason, it is a common assumption in most related works for epidemic spreading on Opportunistic Networks [36,43,73, 76,113]. Furthermore, analyses of real-world traces, suggesting that the exponential distribution can sometimes approximate the distribution of the inter-contact times [25,36], or at least the tail of it [16,68]. This exponential tail is also supported by known results about the hitting times of

[1]We ignore the actual contact duration for simplicity and assume that contacts are instantaneous, since bandwidth concerns are orthogonal to the problem we consider here.

random walks [2]. Finally, findings of two recent analytic studies are consistent with our model: (a) even if the aggregate inter-contact time distribution is non-exponential (as suggested in [18]), individual pair contacts might still be exponential but with different rates [108]; and (b) even if the actual contact times are not Poisson this suffices, under certain conditions, to use a Markov Chain based analytical framework, as a good approximation [112].

The assumption of independence (or even stationarity), while not that well supported by real traces, due to temporal or periodic characteristics in real mobility scenarios [35,117,147], is also necessary for analytical tractability (and any hope for closed form expressions). To our best knowledge, departing from the above assumptions (e.g. maintaining independence but allowing for pareto inter-contact time distributions [12,18]), can only be used for asymptotic, convergence analysis about the message delivery delay of a routing protocol, i.e. if it achieves finite or infinite delay.

The second assumption, the heterogeneity of contact rates between different pairs of nodes, as mentioned above, is motivated by analysis of real traces [25,36,108]. For instance, Passarella et al. [108], shown, using data from real-world social networks, that (i) each person interacts and contacts its friends and acquaintances with higher rate as closer their relationship is, and (ii) the contact rates, between any individual and the other nodes, can be approximated by a distribution (which in our case corresponds to the distribution f_λ). Moreover, it is often quite difficult, in a MSNs context, to know all rates λ_{ij} exactly, or estimates might be rather noisy, which justifies our selection for representing mobility heterogeneity in a probabilistic way (i.e. λ_{ij} are randomly drawn from f_λ).

While this contact class is far from exhaustive and cannot directly capture all types of macroscopic structure often observed in real-world networks (e.g. assortativity and community structure [55, 102]), we can use *any* valid probability density function f_λ to create an infinite range of random contact networks (in contrast to homogeneous models that correspond to only *one* function, i.e. $f_\lambda^{HOM}(\lambda) = \lambda_0 = const.$). Different functions lead to classes of contact processes with very different macroscopic characteristics. For example, large σ_λ^2 values imply that the contact frequencies between different pairs are very heterogeneous, e.g. some pairs will rarely contact each other while others much more often. An f_λ symmetric around μ_λ (e.g. uniform distribution) implies a balanced number of high and low rates, while a right-skewed f_λ (e.g. Pareto) describes a network with most pairs having large intercontact times, but few meeting very frequently. Small μ_λ values could correspond to slow moving nodes, e.g. pedestrians, (or large geographical areas), etc.

For these reasons, we believe the above model strikes a good tradeoff, and as will see, allows us to explore the effect of different social-based characteristics and derive interesting insights, which is the main goal of this work. When possible, we will test these insights against real traces as well, to examine the extent to which departures from the above assumptions affect our conclusions.

2.2.2 Epidemic Spreading

In a contact network $\mathcal{N}$, "messages" might be exchanged between nodes. In the context of MSNs, a message could be a data packet, a file, etc.[2]

[2]In other contexts, the "message" could be a rumour or news in an online social network [80], a virus in a computer network [140], a disease in the physical world [100], etc.

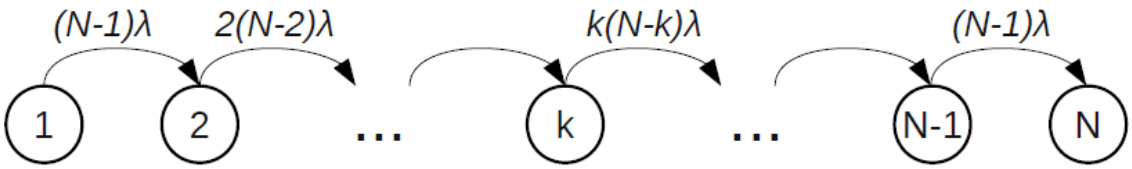


Figure 2.1: Epidemic spreading in a *homogeneous* network with N nodes.

During a contact event, a message currently on one of the nodes *could* be forwarded to ("infect") the other node as well. In the basic epidemic scheme (*epidemic routing* [137]), a message starts from a source node, and a message transfer occurs at *every* contact opportunity involving a node with the message and one without it. To compute the expected message delivery delay of different dissemination mechanisms (e.g. routing protocols, content sharing schemes), we need to split the spreading process in steps, compute the delay of each one of these steps, and use them as the building blocks to calculate the total delay.

In a homogeneous contact network ($\lambda_{ij} = \lambda,\ \forall\{i,j\}$), epidemic spreading can be modelled with a pure-birth Markov chain of N states, as depicted in Fig. 2.1, where a state k denotes the number of "infected" nodes (i.e. nodes with the message). Then, it is easy to show that the *step time* $T_{k,k+1}$ (i.e. the time to move from state k to state $k+1$) is exponentially distributed with rate $k(N-k)\lambda$. Its expected value is then given by $E[T_{k,k+1}] = \frac{1}{k(N-k)\lambda}$, and, therefore, one could straightforwardly calculate the expected spreading time.

However, introducing different contact rates for each pair of nodes complicates the problem. The message dissemination process depends on *which nodes exactly* have the message (in contrast to the homogeneous case, where we only need to track the *number* of infected nodes). As an example, in Fig. 2.2, we present the Markov Chain of a message epidemic spreading in a heterogeneous network with four nodes, $\{A, B, C, D\}$. This Markov Chain is composed of 8 states (or 15 states, if we consider different starting nodes), whereas the respective Markov Chain of an homogeneous network with 4 nodes would be composed of only 4 states. Hence, it becomes evident that the complexity increases quickly, even for this simple 4-node network. In a network with N nodes, the state space explodes with $2^N - 1$ total states, or, equivalently, with $\binom{N}{k}$ different states for *step* k (i.e. when k nodes have the message / are "infected"), and only numerical solutions (and only for N not large) [73] or upper bounds using rough spectral arguments [113], are allowed.

While keeping track of all these states, their probabilities and the rates between them, could be done recursively, the task becomes intractable. To avoid this complication, the main idea behind our results is to prove that, in the limit of large N, all such starting states become *statistically equivalent*, and then collapse them.

2.2.3 Asymptotic Analysis

In this section, we study the *step delay* $T_{k,k+1}$, which is the building block for the calculation of the total epidemic spreading delay, and derive results for its expectation in a large (N) heterogeneous contact network.

As said above, the step delay $T_{k,k+1}$ is the time starting when the k^{th} node just received the message (i.e. *any* k nodes are infected) until the $(k+1)^{th}$ node receives it (i.e. *any* $k+1$ nodes are infected). The calculation of its expectation, $E[T_{k,k+1}]$, involves three sources of randomness:

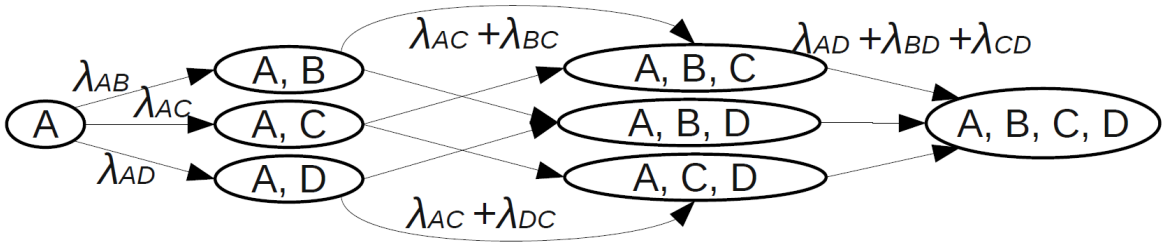


Figure 2.2: Epidemic spreading in a *heterogeneous* network with 4 nodes

(i) A network is initially created according to $f_\lambda(\lambda)$. In other words, $N(N-1)/2$ contact rates λ_{ij} are drawn independently from $f_\lambda(\lambda)$. The resulting graph or (symmetric) contact rate matrix $\mathbf{\Lambda} = \{\lambda_{ij}\}$ is a contact network *instance*[3].

(ii) When at step k, there are k nodes with the message. Conditioned on $\mathbf{\Lambda}$, $T_{k,k+1}$ is a random variable whose distribution will also depend on the actual set of k nodes that have the message, and their contact rates with the remaining nodes. Let $\mathbf{C}_k^m$ denote this set, where m is an integer indicating one of the $\binom{N}{k}$ possible sets of infected relays at step k.

(iii) Finally, conditional on both the network instance $\mathbf{\Lambda}$ and $\mathbf{C}_k^m$, $T_{k,k+1}$ will also depend on the randomness of the inter-contact times involved.

More specifically, let i and j be two nodes, where $i \in \mathbf{C}_k^m$ and $j \notin \mathbf{C}_k^m$, and t_{ij} be the next time they contact after the time the k^{th} node received the message. As the next message exchange will take place when any of the nodes with the message contacts any of the nodes without it, the step delay is given by $T_{k,k+1} = \min_{i \in \mathbf{C}_k^m \; j \notin \mathbf{C}_k^m} \{t_{ij}\}$. Moreover, since t_{ij} are independent, exponentially distributed random variables with rate λ_{ij}, $T_{k,k+1}$ is also exponentially distributed with rate $\sum_{i \in \mathbf{C}_k^m} \sum_{j \notin \mathbf{C}_k^m} \lambda_{ij}$:

$$t_{ij} \sim exp(\lambda_{ij}) \quad \Rightarrow \quad T_{k,k+1} \sim exp\left(\sum_{i \in \mathbf{C}_k^m} \sum_{j \notin \mathbf{C}_k^m} \lambda_{ij} \right) \tag{2.1}$$

and, thus [121]

$$E\left[T_{k,k+1} | \mathbf{C}_k^m\right] = \frac{1}{\sum_{i \in \mathbf{C}_k^m} \sum_{j \notin \mathbf{C}_k^m} \lambda_{ij}} \tag{2.2}$$

Using the properties of conditional expectation, we get the expected delay for the transition

[3]In the following analysis, we will assume that the message is spread in such a network instance with contact rate matrix $\mathbf{\Lambda}$, and, thus, all the expressions will be considered to be conditional on $\mathbf{\Lambda}$. However, for a clearer presentation of the analysis and the results, we will not denote in our expressions the condition on $\mathbf{\Lambda}$.

from step k to step $k+1$:

$$\begin{aligned} E\left[T_{k,k+1}\right] &= \sum_{m=1}^{\binom{N}{k}} E\left[T_{k,k+1}|\mathbf{C}_k^m\right] \cdot P\{\mathbf{C}_k^m\} \\ &= \sum_{m=1}^{\binom{N}{k}} \frac{1}{\sum_{i\in\mathbf{C}_k^m}\sum_{j\notin\mathbf{C}_k^m}\lambda_{ij}} \cdot P\{\mathbf{C}_k^m\} \\ &= \sum_{m=1}^{\binom{N}{k}} \frac{1}{S_k^m} \cdot P\{\mathbf{C}_k^m\} \end{aligned} \tag{2.3}$$

where we denoted

$$S_k^m = \sum_{i\in\mathbf{C}_k^m}\sum_{j\notin\mathbf{C}_k^m}\lambda_{ij} \tag{2.4}$$

The problem in Eq. (2.3) is that keeping track of the probabilities $P\{\mathbf{C}_k^m\}$ is exceedingly complex, and even if we did (e.g. recursively) it would not lead to a useful expression. Instead, we will follow a different approach to compute the expected delay $E[T_{k,k+1}]$:

To derive our main result (Theorem 1) for the expected step delay $E\left[T_{k,k+1}\right]$ (Eq. (2.3)), we will need Lemmas 1 and 2.

Hence, let us first define the random variable S_k as

$$P\{S_k = S_k^m\} = P\{\mathbf{C}_k^m\} \tag{2.5}$$

and the random variable X_k as $X_k = \frac{S_k}{k(N-k)}$, i.e.

$$P\left\{X_k = \frac{S_k^m}{k(N-k)}\right\} = P\{\mathbf{C}_k^m\} \tag{2.6}$$

Now we can state Lemma 1 that gives the first two moments of the random variable S_k, and Lemma 2 that shows how the random variable X_k converges as the network size N increases. The proofs of Lemmas 1 and 2 can be found in Appendices 2.6.1 and 2.6.2, respectively.

Lemma 1. *The expectation and variance of the random variable S_k at step k, are given by*

$$\begin{aligned} E[S_k] &= k(N-k)\cdot\mu_\lambda\cdot(1-\epsilon_k) \\ Var\left[S_k\right] &= k(N-k)\cdot\sigma_\lambda^2\cdot(1-\delta_k) \end{aligned}$$

where $\epsilon_k = O\left(\frac{\lambda_{max}}{N}\right)$ and $|\delta_k| = O\left(\frac{\lambda_{max}^2}{N}\right)$.

Lemma 2. *As the network size N increases, the random variable X_k converges as follows*

$$X_k \xrightarrow{m.s.} \mu_\lambda$$

where $\xrightarrow{m.s.}$ denotes convergence in mean square.

Using the above Lemmas, we can prove Theorem 1, which suggests that in a large Heterogeneous Contact Network the expected step delay at a step k, can be approximated with infinite accuracy as following.

$$E[T_{k,k+1}] \approx \frac{1}{k(N-k)\mu_\lambda}$$

Theorem 1. *As the network size N increases, the relative error RE_k between the expected step delay $E[T_{k,k+1}]$ and the quantity $\frac{1}{k(N-k)\mu_\lambda}$ converges to zero*

$$\lim_{N\to\infty} RE_k = \lim_{N\to\infty} \frac{E[T_{k,k+1}] - \frac{1}{k(N-k)\mu_\lambda}}{E[T_{k,k+1}]} = 0$$

Proof. Lemma 2 shows the convergence in mean square for X_k. Therefore, it follows directly that X_k converges in probability as well [70, p. 140-141]

$$X_k \xrightarrow{m.s.} \mu_\lambda \quad \Rightarrow \quad X_k \xrightarrow{p} \mu_\lambda \tag{2.7}$$

where $\xrightarrow{p}$ denotes convergence in probability.

Let us, now, define the random variable Y_k as $Y_k = \frac{1}{X_k} = \frac{k(N-k)}{S_k}$, with probability distribution

$$P\left\{Y_k = \frac{k(N-k)}{S_k^m}\right\} = P\{\mathbf{C}_k^m\} \tag{2.8}$$

Since (see Eq. (2.7)) $X_k \xrightarrow{p} \mu_\lambda$, it also holds that [70, Thm. 5.23, p. 148]

$$Y_k = \frac{1}{X_k} \xrightarrow{p} \frac{1}{\mu_\lambda} \tag{2.9}$$

Moreover, since each contact rate λ_{ij} takes values in the interval $[\lambda_{min}, \lambda_{max}]$, it is easy to see that

$$\frac{1}{\lambda_{max}} \leq Y_k \leq \frac{1}{\lambda_{min}} \tag{2.10}$$

Using Eq. (2.10) and the definition of *uniform integrability* [70, Def. 5.15, p. 142], it follows that Y_k is uniformly integrable $\forall N$ and $\forall k \in [1, N-1]$, i.e.

$$\lim_{\alpha\to\infty} \sup_N E\left[|Y_k|; \{|Y_k| > \alpha\}\right] = 0 \tag{2.11}$$

because $P\{|Y_k| > \alpha\} = 0$ for $\alpha > \frac{1}{\lambda_{min}}$.

Eq. (2.9) states that Y_k converges in probability to $\frac{1}{\mu_\lambda}$ and Eq. (2.11) that Y_k is uniformly integrable. Therefore, [70, Thm. 5.17, p. 144], it follows that Y_k converges in mean value (denoted with $\xrightarrow{m.}$) to $\frac{1}{\mu_\lambda}$:

$$Y_k \xrightarrow{m.} \frac{1}{\mu_\lambda} \quad \text{or} \quad E[Y_k] = E\left[\frac{1}{X_k}\right] \to \frac{1}{\mu_\lambda} \tag{2.12}$$

Finally, the relative error RE_k can be written as

$$RE_k = \frac{E[T_{k,k+1}] - \frac{1}{k(N-k)\mu_\lambda}}{E[T_{k,k+1}]} = \frac{E\left[\frac{1}{S_k}\right] - \frac{1}{k(N-k)\mu_\lambda}}{E\left[\frac{1}{S_k}\right]} = \frac{E\left[\frac{1}{X_k}\right] - \frac{1}{\mu_\lambda}}{E\left[\frac{1}{X_k}\right]} = \frac{E[Y_k] - \frac{1}{\mu_\lambda}}{E[Y_k]} \tag{2.13}$$

Taking the limit in Eq. (2.13) for $N \to \infty$ and using Eq. (2.12), gives

$$\lim_{N\to\infty} RE_k = \frac{\frac{1}{\mu_\lambda} - \frac{1}{\mu_\lambda}}{\frac{1}{\mu_\lambda}} = 0 \tag{2.14}$$

□

Table 2.1: Relative Step Delay Error RE_k: Averaged over All Steps and over 100 Network Instances

	$N = 20$	$N = 50$	$N = 100$	$N = 200$	$N = 500$
$CV_\lambda = 0.5$	4.3%	2.8%	2.7%	2.6%	2.5%
$CV_\lambda = 1$	10.6%	4.2%	3.1%	2.7%	2.6%
$CV_\lambda = 1.5$	22.4%	8.2%	4.6%	3.2%	2.6%
$CV_\lambda = 3$	126.7%	34.1%	15.3%	8.2%	3.8%

To verify our results, we test them against simulations. We use a simulator that creates network instances belonging to the class of Heterogeneous Contact Networks (Def. 3), we ran Monte Carlo simulations of epidemic spreading, and calculate the mean step delay (further details for the simulation methodology are given in Section 2.2.5.1). In Table 2.1, we present the values for the relative error RE_k (Theorem 1) in simulation scenarios of different network sizes N and contact rates heterogeneity CV_λ. The values in Table 2.1 correspond to the relative error RE_k averaged over all the steps k of the epidemic process and over 100 different network instances $\mathbf{\Lambda}$ with equivalent characteristics (N, f_λ). It can be seen that in networks with higher heterogeneity CV_λ (and, thus, larger ranges $[\lambda_{min}, \lambda_{max}]$, since we set the mean rate $\mu_\lambda = 1$) the errors are larger, as our theory predicts. However, as the network size increases, the errors for all scenarios become very small.

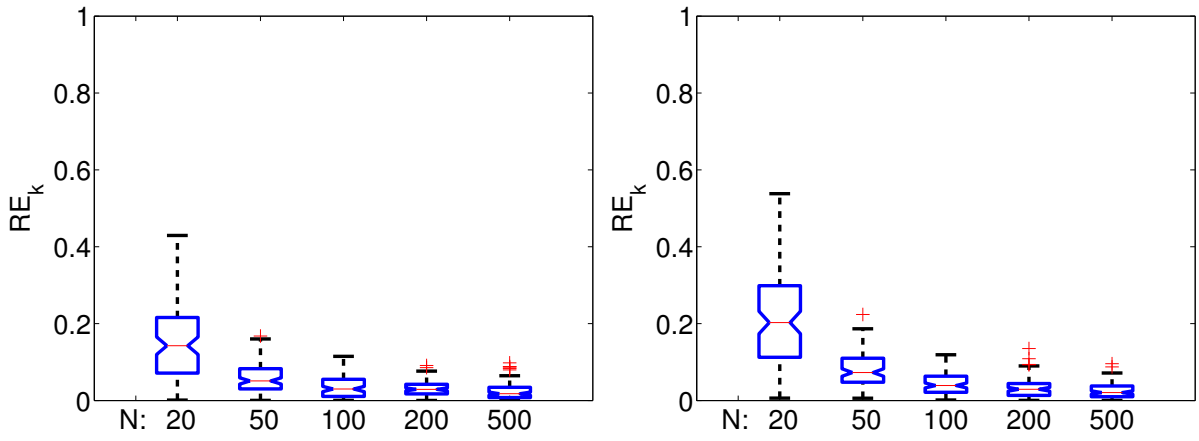
The decrease of the relative errors can be observed also in Fig. 2.3, where we present the distribution (boxplots[4]) of the values of RE_k over the different network instances. Here, the relative errors do not correspond to averaged (over different steps) values, but we present the RE_k at the steps that correspond at the 20% (e.g. in the scenario with $N = 100$, we present the relative errors in the step $k = 20$) and 70% of the spreading process, in Fig. 2.3(a) and Fig. 2.3(b), respectively. It can be seen that in later steps the error is slightly larger, which is expected, due to the accumulation of errors from all previous steps. Nevertheless, for large network sizes, the error diminishes for every step considered.

2.2.4 Finite Size Networks

The asymptotic analysis and results of the previous section can be used to predict accurately the spreading delay in large networks. In addition to this, they provide useful insights and guidelines for the analysis of finite cases (small networks). In that sense, in this section, we interpret the results of Section 2.2.3 and derive simple, closed-form approximations for the behavior of *finite size* networks.

Specifically, from Theorem 1, the quantity $\frac{1}{k(N-k)\mu_\lambda}$ can be used as a predictor for the step delay, and the prediction error converges to 0 as networks get larger. For finite cases though, this error might not be negligible. This motivates us to investigate how the approximation for the step delay can be improved (e.g. by adding *higher order terms*). To this end, we consider the following analysis.

[4]In each box, the central horizontal (red) line is the median, the edges of the (blue) box are the 25^{th} and 75^{th} percentiles, the (black) whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as (red) crosses.



(a) spreading step: 20% (b) spreading step: 70%

Figure 2.3: *Relative Step Error* for the step (a) $k = 0.2 \cdot N$ (i.e. message spreading at 20% of the network) and (b) $k = 0.7 \cdot N$. Each boxplot corresponds to a different network size N (with $\mu_\lambda = 1$ and $CV_\lambda = 1.5$). Box-plots show the distribution of the Relative Step Error RE_k for 100 different network instances of the same size.

At first, from Eq. (2.3) we can express the mean step time $E[T_{k,k+1}]$ as

$$E\left[T_{k,k+1}\right] = \sum_{m=1}^{\binom{N}{k}} \frac{1}{S_k^m} \cdot P\{\mathbf{C}_k^m\} = E\left[\frac{1}{S_k}\right] \tag{2.15}$$

where S_k is defined in Eq. (2.5). Since we do not know the probabilities $P\{\mathbf{C}_k^m\}$ (i.e. the exact distribution of S_k), it is not possible to calculate the quantity $E\left[\frac{1}{S_k}\right]$. However, $E\left[\frac{1}{S_k}\right]$ is the expectation of a function of S_k (i.e. the function $g(x) = x^{-1}$), and thus we can approximate it by using the *Delta method* [103], where the expectation of a function of a random variable (i.e. $E[g(S_k)] \equiv E\left[\frac{1}{S_k}\right]$) is approximated using the Taylor expansion of the function and the first moments of the random variable (i.e. $E[S_k]$, $Var\left[S_k\right]$, etc.).

The calculation of these moments though, still depends on the knowledge of the probabilities $P\{\mathbf{C}_k^m\}$, and exact expressions cannot be found. Hence, to proceed further and be able to derive useful results, we approximate the first two central moments of S_k, by neglecting the terms ϵ_k and δ_k in the expressions of Lemma 1, i.e.

$$E[S_k] \approx k(N-k) \cdot \mu_\lambda \tag{2.16}$$

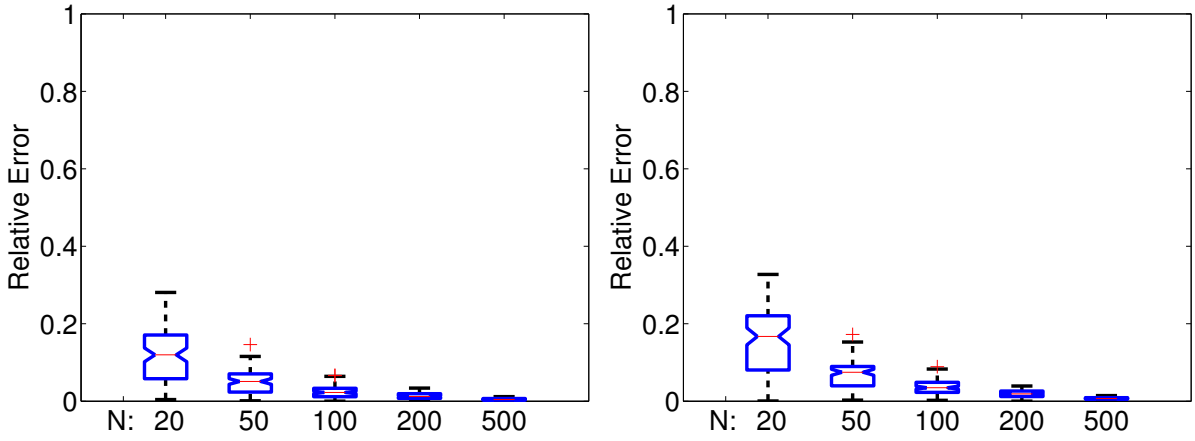
$$Var[S_k] \approx k(N-k) \cdot \sigma_\lambda^2 \tag{2.17}$$

These approximations, as Lemma 1 implies, become more accurate as (i) the size of the network increases, or (ii) the heterogeneity of the contact rates decreases. To further support this argument, we present some initial simulation results. Table 2.2 and Fig. 2.4 (in a similar way to the previous section) show the relative errors between the quantity $E[X_k]$ and the approximation we consider, $k(N-k)\mu_\lambda$. As it can be seen, the approximation is relatively accurate even for moderate network sizes.

Now, using the *Delta method* and the expressions of Eq. (2.16) and Eq. (2.17), we provide in Result 1 a *second order* approximation for the expected step delay.

Table 2.2: Relative Error $\frac{E[X_k]-k(N-k)\mu_\lambda}{k(N-k)\mu_\lambda}$: Averaged over All Steps and over 100 Network Instances

	$N=20$	$N=50$	$N=100$	$N=200$	$N=500$
$CV_\lambda = 0.5$	3.3%	1.3%	0.7%	0.3%	0.1%
$CV_\lambda = 1$	8.3%	3.0%	1.7%	0.9%	0.3%
$CV_\lambda = 1.5$	15.3%	6.6%	3.5%	1.9%	0.7%
$CV_\lambda = 3$	38.7%	21.6%	12.1%	7.1%	2.8%



(a) spreading step: 20%

(b) spreading step: 70%

Figure 2.4: *Relative Error* for the step (a) $k = 0.2 \cdot N$ (i.e. message spreading at 20% of the network) and (b) $k = 0.7 \cdot N$. Each boxplot correspond to a different network size N (with $\mu_\lambda = 1$ and $CV_\lambda = 1.5$). Box-plots show the distribution of the Relative Error $\frac{E[X_k]-k(N-k)\mu_\lambda}{k(N-k)\mu_\lambda}$ for 100 different network instances of the same size.

Result 1. *In a Heterogeneous Contact Network (Def. 3) the expected step delay can be approximated by*

$$E[T_{k,k+1}] = \frac{1}{k(N-k)\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{k(N-k)}\right) \tag{2.18}$$

Proof. To estimate $E\left[\frac{1}{S_k}\right] = E[g(S_k)]$, at first we express the function $g(S_k) = \frac{1}{S_k}$ as a Taylor series expansion, centered at $E[S_k]$, the mean value of S_k.

$$T_g(S_k) = \sum_{n=0}^{\infty} \frac{g^{(n)}(E[S_k])}{n!}(S_k - E[S_k])^n = \sum_{n=0}^{\infty} \frac{(-1)^n (S_k - E[S_k])^n}{(E[S_k])^{n+1}} \tag{2.19}$$

We can approximate $g(S_k)$ by taking the first m terms of the Taylor series. That will result in:

$$g(S_k) \approx \sum_{n=0}^{m} \frac{(-1)^n}{(E[S_k])^{n+1}}(S_k - E[S_k])^n \tag{2.20}$$

An approximation for the mean value of $g(S_k)$ follows after taking the expectation of both sides in the last equation.

$$E[g(S_k)] \approx \sum_{n=0}^{m} \frac{(-1)^n}{(E[S_k])^{n+1}} M_n \tag{2.21}$$

where $M_n = E[(S_k - E[S_k])^n]$ is the n^{th} central moment.

This method, of approximating a function with a finite Taylor sum and taking the expectation of it for evaluating the mean value of the function, is widely known as the *delta method* [27,103].

Considering $m = 2$ in Eq. (2.21) and using the expressions of Eq. (2.16) and Eq. (2.17) for the moments $M_0 = E[S_k]$ and $M_2 = Var[S_k]$, proves the result. □

Remark: In the Delta method, different number of terms of the Taylor series can be taken into account, depending on the required accuracy (the more terms one considers, the more accurate the result). For example, taking only the first term ($m = 0$), we get the asymptotic expression, i.e. $E[T_{k,k+1}] = \frac{1}{k(N-k)\mu_\lambda}$. As a better approximation, we consider here the first three terms ($m = 2$) of the Taylor series, which involve the first two moments of S_k. Our choice for using the approximation that depends on the first two moments is a trade off between usability and expressibility of the result, and its accuracy.

2.2.5 Delivery Delay of Opportunistic Routing Protocols

Having found the necessary approximations for individual epidemic steps in heterogeneous scenarios, we turn our attention to applications of these results. Specifically, we use the basic building blocks of our analysis (i.e. step delay $T_{k,k+1}$) to predict the end-to-end delivery delay for three routing protocols, namely, *epidemic* routing [137], *2-hop* routing [43] and *Spray and Wait* routing [129].

We briefly present here the mechanism of these schemes and how the expected delivery delay can be computed for each of them. Table 2.3 gives the approximative closed-form expressions for the delivery delay (corresponding to the approximation of Result 1 for the step delay), while detailed derivations of the formulas can be found in Section 2.6.3.

Epidemic routing

In *unicast* epidemic routing, we are interested in the time until a given destination node receives the message. Assuming the selection of the source and destination nodes is random, the probability that the destination node is the k^{th} node to receive the message is the same for every k, i.e. $\frac{1}{N-1}$. Consequently, adding up the expected delays of all steps (due to the linearity of expectation) and multiplying them with the probabilities $\frac{1}{N-1}$, we get

$$E[T_{epid}^{uni}] = \frac{1}{N-1}\sum_{n=2}^{N}\sum_{k=1}^{n-1} E[T_{k,k+1}] = \frac{1}{N-1}\sum_{k=1}^{N-1}(N-k)E[T_{k,k+1}]. \tag{2.22}$$

where $E[T_{k,k+1}]$ can be calculated e.g. by the approximate expression of Result 1.

2-hop routing

In the 2-hop routing scheme, the source sends the message to every node it meets, like in epidemic routing. However, other nodes receiving the message can only give it directly to the destination, when and if they encounter it.

Therefore, in *step* k (the source and $k-1$ relays carry the message) there are $N-1$ possible meeting events in which a message exchange can take place, i.e. (i) $N-k-1$ possible meetings between the source and a node without the message, other than the destination, and (ii) k possible meetings between the relay nodes (including the source) and the destination. Due to randomness, the probability that the destination node will be involved in the exact next meeting event with message exchange is $\frac{k}{(N-k-1)+k} = \frac{k}{N-1}$. Based on the previous observations, we can at first compute the probability the message to be delivered at each step k and the expected step delay ($E[T_{k,k+1}^{2-hop}]$), from which we can eventually derive the expected delivery delay.

Spray and Wait (SnW) routing

In the Spray and Wait scheme[5], the source generates L copies of the message and when it meets another node, it gives to it half of the messages it holds at that time (if it holds more than one). The same mechanism applies when a relay node with more than one copies meets another node without the message. Eventually there would be L nodes (including the source) holding the message. If the message is not delivered to the destination before the L message copies are spread (*spray phase*), it will be delivered the first time any of the L nodes with the message meets the destination (*wait phase*).

The expression for the delivery delay is calculated by following a similar procedure as in the *2-hop* routing case.

2.2.5.1 Model Validation

Synthetic Scenarios

In order to validate the accuracy of our predictions for the message delivery delay under different routing schemes, we first compare them against simulations of various *synthetic* scenarios

[5]Here we describe the *binary* Spray and Wait, which is the scheme with the lowest expected delivery delay among all the SnW-based schemes (e.g. *source SnW*) [129].

Table 2.3: approximative expressions for the Expected Delivery Delay of different routing protocols.

Epidemic	$E[T_D^{(epid)}] \approx \frac{1}{N \cdot \mu_\lambda} \cdot \left(\ln(N) + CV_\lambda^2 \cdot \frac{1.65 \cdot N + 2 \cdot \ln(N)}{N^2} \right)$
2-hop	$E[T_D^{(2-hop)}] = A_{N-1} \cdot \sum_{k=1}^{N-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+1} \cdot (N-k-1)!} \approx \frac{\sqrt{\frac{\pi}{2}}}{\sqrt{N} \cdot \mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N} \right)$
SnW, L copies	$E[T_D^{(SnW)}] \leq A_{N-1} \cdot \sum_{k=1}^{L-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+1} \cdot (N-k-1)!} + (L \cdot A_{N-1} + A_L) \cdot \frac{(N-1)!}{(N-1)^L \cdot (N-L-1)!}$

$$\text{where } A_m = \frac{1}{m\mu_\lambda} \cdot \left[1 + \frac{CV_\lambda^2}{m} \right]$$

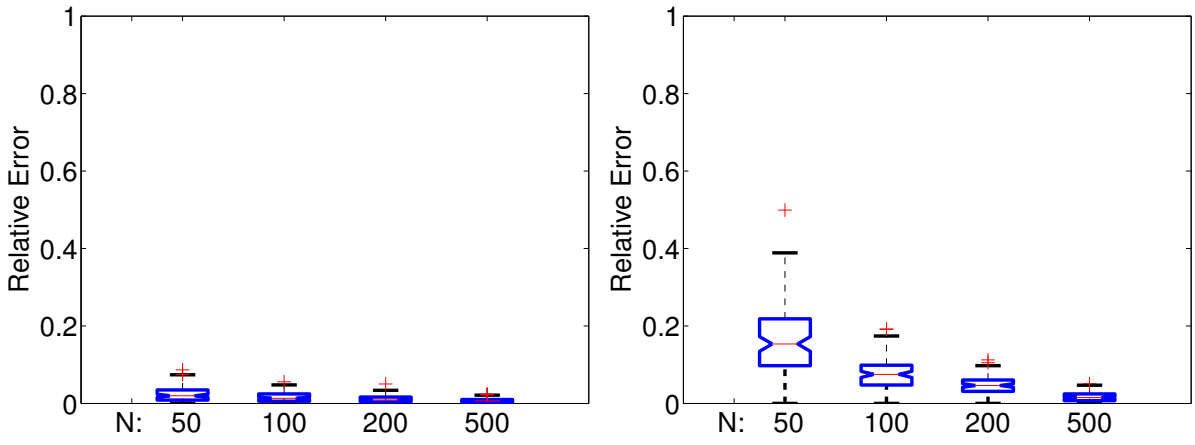
belonging to the class of Heterogeneous Contact Networks (Def. 3). We use Monte Carlo simulations to examine the accuracy of our various analytical expressions (i) in finite size networks and (ii) as a function of other parameters of interest (e.g. statistics of the contact rates generating function f_λ).

In each simulation, we create a network of N nodes and a contact pattern by generating a $N \times N$ matrix $\mathbf{\Lambda} = \{\lambda_{ij}\}$. Each entry λ_{ij} characterizes the contact process of the pair of nodes i and j, and it takes values drawn from a chosen distribution f_λ with mean μ_λ and variance σ_λ^2 ($CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$). Then for each pair we generate a sequence of contact events with exponentially distributed intercontact times with rate λ_{ij}.

For every network instance $\mathbf{\Lambda}$, we run 1000 message spreading simulations, choosing randomly the source and destination nodes, and calculate the average delivery delay. We have considered scenarios with contact rate distributions f_λ with varying heterogeneity (CV_λ). Without loss of generality and for a clearer comparison we set the average contact rate equal to the unit, i.e. $\mu_\lambda = 1$.

Fig. 2.5 shows the relative error between the expected delivery delay of epidemic routing (for different network instances) and the corresponding theoretical prediction (Table 2.3) under two set of scenarios with different contact rates heterogeneity. It can be seen that the accuracy of our prediction is significant, even for small networks, when the heterogeneity is not high (Fig. 2.5(a)). Although the accuracy decreases with heterogeneity (Fig. 2.5(a)), for networks larger than a hundred nodes, the relative errors are less than 10%.

Similar observations can be made also in Fig. 2.6 for the expected delivery delay of Epidemic and Spray and Wait routing. In Fig. 2.6(a), we present simulation results in four scenarios with different network sizes (N) and mobility heterogeneity (CV_λ), where epidemic routing is used for delivering messages. Fig. 2.6(b) shows similar results for the case of Spray and Wait routing. Networks with $N = 500$ nodes and varying mobility heterogeneity are considered. In all scenarios, simulation results are averaged over 100 network instances. As it can be seen, the



(a) $\mu_\lambda = 1$, $CV_\lambda = 1$ (b) $\mu_\lambda = 1$, $CV_\lambda = 3$

Figure 2.5: Relative Error between the *simulated expected delivery delay of epidemic routing* and the *theoretical approximation*. Each boxplot corresponds to a different network size N. Box-plots show the distribution of the Relative Error for 100 different network instances of the same size.

predictions of our approximate expressions are quite close to the simulated values.

Real Mobility Traces

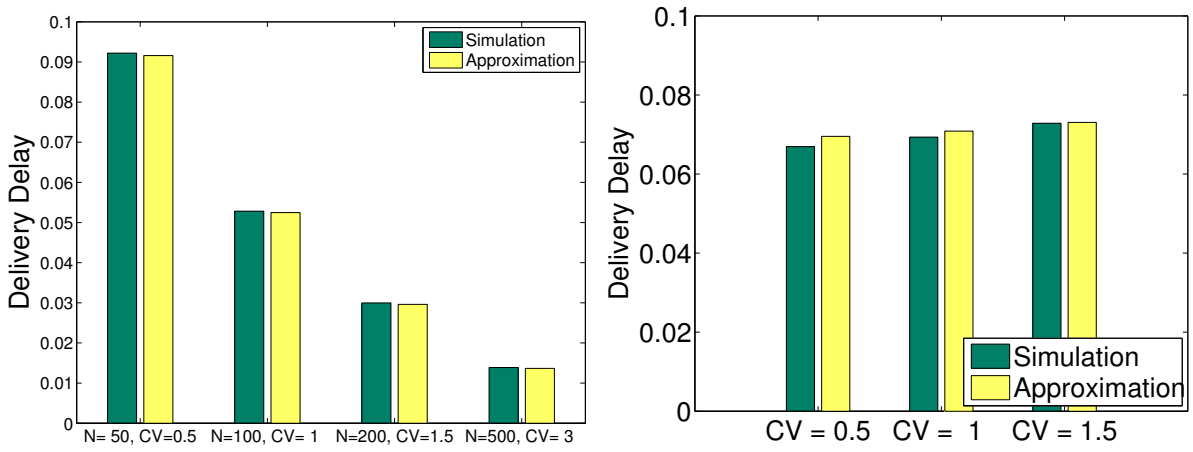
The above simulation results show that our analytical predictions achieve significant accuracy even in finite networks whose mobility patterns fall under the class of Heterogeneous Contact Networks. While these contact classes are rather broad, whether they capture "real" scenarios, and to what extent, depends on the application setting, contact scenario, etc.

In the context of MSNs, some mobility traces collected in real experiments and/or networks do exist. Arguably, the size of most of them is small and they represent each only a single instance of the random mobility process at play, often with a number of measurement complications and errors. Nevertheless, it is of interest to see how our performance predictors behave in some of these scenarios, and whether they can capture the quantities of interest (even if qualitatively), despite the considerably higher complexity (e.g. community structure) of such scenarios, and departures from the assumptions for which our predictors are designed.

To this end, we use the following sets of real mobility traces: (i) *Cabspotting* [118], which contains GPS coordinates from 536 taxi cabs collected over 30 days in San Francisco, and (ii) *Infocom* [125], which contains traces of Bluetooth sightings of 78 mobile nodes from the 4 days iMotes experiment during Infocom 2006. We also generated mobility traces with two recent mobility models that have been shown to capture well different aspects of real mobility traces, namely, TVCM [57] and SLAW [77]. In order to compare with analysis, we parse each trace and estimate the mean contact rate for all pairs $\{i, j\}$. We then produce estimates for the 1^{st} and 2^{nd} moments of these rates, $\hat{\mu}_\lambda$ and $\hat{\sigma}^2_\lambda$, and use them in our analytical expressions.

Fig. 2.7 shows the message delay under epidemic and 2-hop routing. Source and destination are chosen randomly in different runs and messages are generated in random points of the trace.

The first thing to observe is that delay values span a wide range of values for different source-destination pairs. This implies a large amount of heterogeneity in the "reachability" of different nodes. Our analytical predictions are shown as thick dark horizontal lines. As it can



(a) Epidemic routing

(b) Spray and Wait

Figure 2.6: Delivery Delay of (a) epidemic and (b) spray and wait routing in different scenarios of Heterogeneous Contact Networks. Simulation results are averaged over 100 network instances.

be seen, our result is in most cases close to the median and in almost all cases between the 25^{th} and 75^{th} percentile of the delay observed in both the real traces and mobility models[6]. It is somewhat remarkable that our delay predictors are close to the actual results (qualitatively or even quantitatively in some cases) in a range of real or realistic scenarios; studies of these scenarios reveal considerable differences to the much simpler contact classes for which our results are derived. We should also be careful not to jump to generalizations about the accuracy of these results in all real scenarios, as we are aware of situations that could force our predictors to err significantly. Nevertheless, we believe these results are quite promising in the direction of finding simple, usable analytical expressions even for complex, heterogeneous contact scenarios.

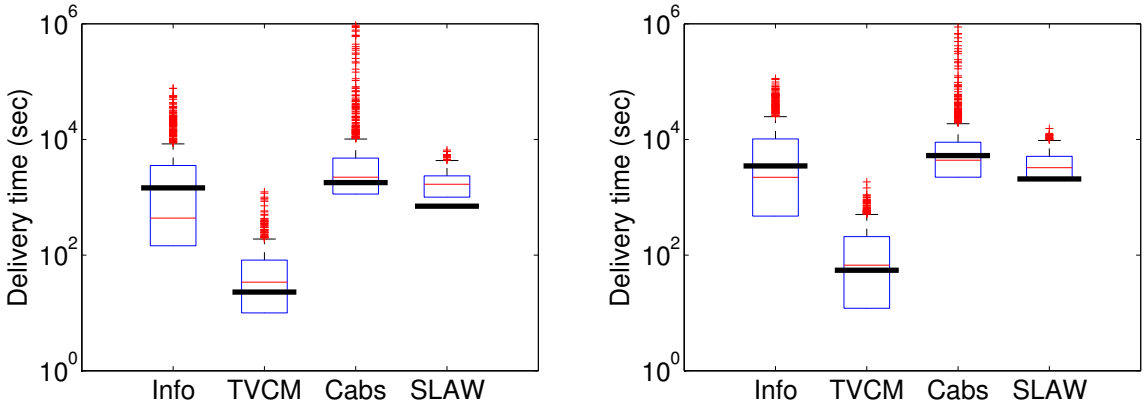
2.3 Sparse Contact Graphs

After having investigated to what extend the heterogeneity of contact rates affects the delay of information spreading, we now remove the second unrealistic key assumption. Specifically, the network contact model of Def. 3 assumes that every pair of nodes meets with non-zero rate. However, the network contact graph is not necessarily fully mixed (Fig. 2.8(a)) but can be sparse and/or very heterogeneous (Fig. 2.8(b)). To describe such sparse networks, we propose the following two models for the network contact graph. These models are (i) based on random graphs and thus the network can be studied using analytic methods, and (ii) can describe networks arbitrarily sparse (Poisson Contact Graph model - Section 2.3.1) and with arbitrarily heterogeneous nodes (Configuration Contact Graph model - Section 2.3.2).

2.3.1 Poisson Contact Graph

We first extend the Heterogeneous Contact Network model of Def. 3, by allowing some pairs to never meet, i.e. we allow $\lambda_{ij} = 0$. Specifically, we consider the following class of Contact

[6]We tend to underestimate the delay in SLAW, a mobility model that was designed to capture power-law characteristics of contact meetings [77]



(a) epidemic routing

(b) 2-hop routing

Figure 2.7: Box-plots of the unicast (i.e. message delivery) delay under epidemic and 2-hop routing. On each box, the central horizontal line is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually as crosses. The thick lines represent the theoretical values predicted by our model.

Networks

Definition 4 (Heterogeneous Poisson Contact Network)**.** *For each pair of nodes i and j the following holds: (i) with probability $1-p_s$ they* never *contact each other, (ii) with probability p_s they contact with rate λ_{ij}, according to a contact process as defined in Def. 3.*

In other words, we now first create a *Poisson (or Erdös-Renyi) graph* [102] between nodes. We then assign rates λ_{ij}, as before, but only to the existing links. With the parameter p_s, we can now also capture arbitrarily sparse scenarios, where each node meets only a percentage of all nodes[7].

The following corollary suggests that the previous analysis (Section 2.2) is valid also in the case of a sparse network (Def. 4) and the results hold by just modifying the values of μ_λ and σ^2_λ.

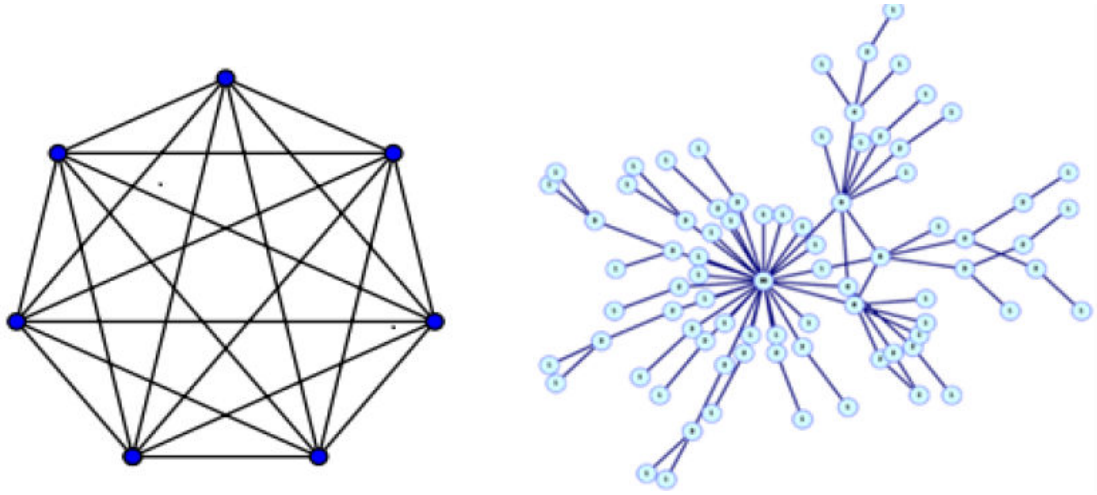
Corollary 1. *Under a Heterogeneous Poisson Contact Network (Def. 4), the theoretical results for a Heterogeneous Contact Network (Def. 3), are modified by substituting the moments of the contact rate distribution (μ_λ and σ^2_λ) with the expressions*

$$\mu_{\lambda(p)} = p_s \cdot \mu_\lambda$$
$$\sigma^2_{\lambda(p)} = p_s \cdot \left[\sigma^2_\lambda + \mu^2_\lambda \cdot (1-p_s)\right]$$

Corollary 1 can be proved similarly to the theoretical results of Section 2.2. In Appendix 2.6.4, we present a sketch of this proof comprising the main analytical arguments and differences compared to the analysis for the full-meshed network case.

To validate Corollary 1, we perform simulations as in the previous section. In Fig. 2.9 we compare our theoretical predictions (calculated from the expressions of Table 2.3, which we

[7]We *do* assume that the probability p_s is large enough for connectivity to be achieved. In practice, the theory of Poisson graphs tells us that connectivity can be achieved with arbitrary low p_s as long as N is large enough (with percolation occurring at an average degree as low as 1 in the limit) [102].



(a) Full mesh graph

(b) Sparse graph

Figure 2.8: Representation of contact networks with graphs

modified according to Corollary 1) against simulation results for the expected delivery delay of Epidemic and Spray and Wait routing. It can be seen the predictions are not far from the simulation results in these sparse network cases. However, they are less accurate than the corresponding dense networks of the same size (see Fig. 2.6). This is due to the fact that the poisson contact graph introduces further randomness and, thus, more diversity in the spreading process. Hence, the same accuracy is achieved for larger network sizes.

2.3.2 Configuration Model Contact Graph

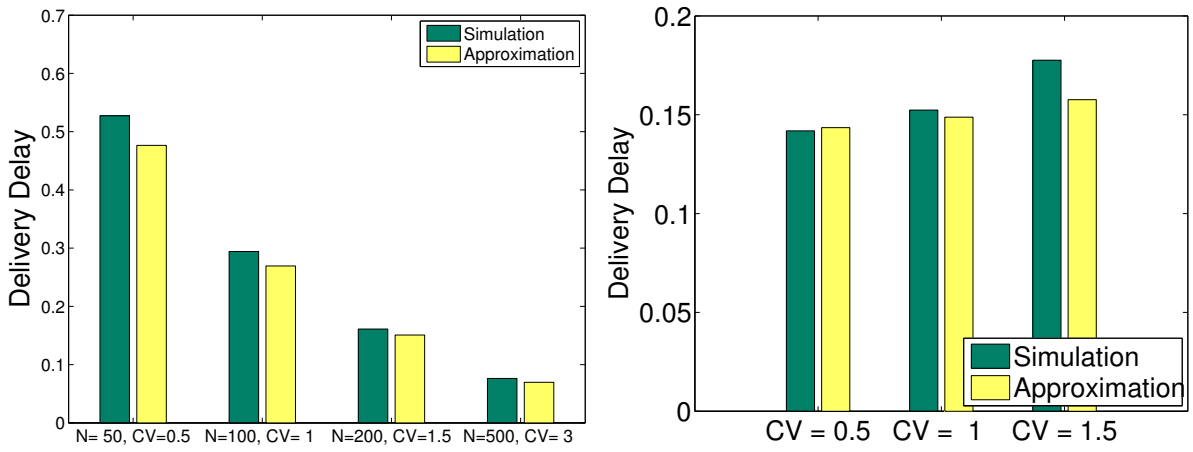
As presented in the previous section, modeling the network's contact graph with a Poisson graph allows us to capture different levels of sparseness by selecting appropriately the probability p_s. However, contact graphs of real networks, in general, have more complex characteristics than a Poisson graph. In particular, the Poisson distribution cannot always approximate accurately the nodes degree distribution[8] [102].

To this end, in this section, we add further complexity and heterogeneity in the network contact graph model (and thus, better approximate real scenarios), by allowing different nodes to have different degrees. Therefore, we would be able to describe networks where some nodes can meet/contact a lot of nodes, e.g. because the are more mobile or because they visit more crowded areas, whereas others meet only a few nodes.

To incorporate such contact graphs in our analysis, we use the Configuration Model [95,102], which creates random graphs that can have any generic degree distribution, and, thus, it can capture the degree characteristics of real-world scenarios and networks (Fig. 2.8(b)).

Definition 5 (Configuration Model)**.** *Given a network size N and a degree distribution $\boldsymbol{p_d}$ (or a degree sequence d_i, $i = 1, ..., N$), the Configuration Model draws random instances among all the graphs $\mathcal{G}$, with N vertices, for which the degree distribution is $\boldsymbol{p_d}$. Connections between nodes*

[8]The degree of a node/vertex is the number of edges connected to it, or, in the context of MSNs, the number of other nodes it *ever* meets.



(a) Epidemic routing (b) Poisson Mixing Network

Figure 2.9: Delivery Delay of (a) epidemic and (b) spray and wait routing in different scenarios of Heterogeneous Poisson Contact Networks with ($p_s = 0.2$). Simulation results are averaged over 100 network instances.

are made randomly, and the probability of having a link between two nodes i and j is proportional only to the degrees of i and j.

The main strengths of the Configuration Model are that: (i) It can describe networks in which the degrees of the vertices can follow any arbitrary distribution. The degree distribution of the vertices[9] is an important characteristic of contact networks and it can determine the evolution of processes on the network (e.g. whether information, a virus, or a disease manages to spread.) [102]; (ii) It is based on random graphs and thus the network can be studied using analytic methods, which is the goal of our work.

Summarizing, in this section we consider the following class of Contact Networks

Definition 6 (Heterogeneous Configuration Model Contact Network)**.**
– Given a degree distribution $\boldsymbol{p_d}$, with mean value μ_d and variance σ_d^2 (and $CV_d = \frac{\sigma_d}{\mu_d}$), a contact graph $\mathcal{G}$ is generated by the Configuration Model.
– Each pair of nodes i and j, connected with an edge, contact each other with rate λ (equal for all contacting pairs), according to a contact process as defined in Def. 3.

With the above contact class, we are able to capture more complex characteristics, with respect to contact graph structure, than the classes of Def. 3 and 4. However, assuming both heterogeneous node degrees (i.e. configuration model) and heterogeneous contact rates (different λ_{ij} for each pair $\{i, j\}$) would make the problem of investigating the effects on the communication performance analytically intractable. To this end, in our analysis, we assume homogeneous contact rates λ (for nodes that do contact each other), which allows us to derive useful, closed form results, and then we test this approximation with simulations on networks with heterogeneous rates λ_{ij} as well (Section 2.3.2.2).

[9]We will use the terms *vertex* and *node* interchangeably.

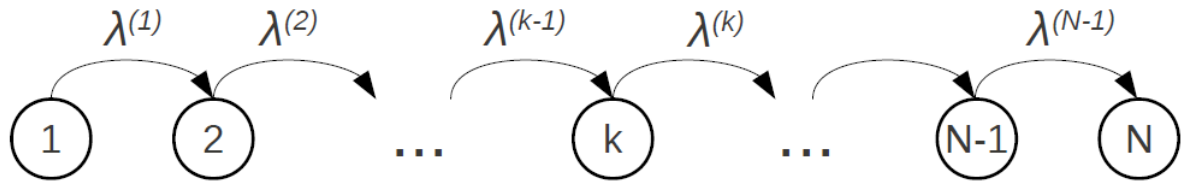


Figure 2.10: Epidemic spreading over a Heterogeneous Configuration Model Contact Network with N nodes. The transition rate from state k to state $k+1$ is $\lambda^{(k)}$.

2.3.2.1 Analysis

We consider now, similarly to Section 2.2.2, an epidemic spreading of message over a contact network $\mathcal{N}$ consisting of N total nodes, and we are interested to compute the expected message delivery delay of different dissemination schemes. As before, we define the spreading process to be at *state* k, $k = 1, ..., N-1$ when k nodes have the message (as shown in Fig. 2.10), we denote the set of the "infected" nodes as $\mathbf{C}(k)$, and we refer to the transition from state k to state $k+1$ as *step* k.

Due to the memoryless property of the Poisson contact events, the duration of step k only depends on the sum of contact rates between nodes with the message ($\in \mathbf{C}(k)$) and nodes that have not received it yet ($\notin \mathbf{C}(k)$). In Fig. 2.10, the sum of these rates is denoted as $\lambda^{(k)}$. In a Heterogeneous Configuration Model Contact Network (Def. 6), the contact rates have the same value λ for all node pairs. Hence, $\lambda^{(k)}$ is given by

$$\lambda^{(k)} = \lambda \cdot D^{out}(k) = \lambda \cdot \sum_{i \in \mathbf{C}(k)} \sum_{j \notin \mathbf{C}(k)} \mathbb{I}_{ij} \tag{2.23}$$

where $\mathbb{I}_{ij} = 1$ iff there exists an edge between nodes $i - j$ (i.e. i and j contact each other). $D^{out}(k) = \sum_{i \in \mathbf{C}(k)} \sum_{j \notin \mathbf{C}(k)} \mathbb{I}_{ij}$ is defined as the *out degree* of step k. In other words, the *out degree* is the number of all the possible ways that the message can infect one additional node, when at state k.

Knowing $D^{out}(k)$ (i.e. the number of $i - j$ node pairs that could further spread the message at step k) is enough to derive the total delay of each step. However, in a Heterogeneous Configuration Model Contact Network, $D^{out}(k)$ is a random variable which depends on the degrees of the k nodes that *happen* to get infected first, as shown in Fig. 2.11. What is more, unlike uniform degree models, not all nodes here have the same probability of being infected first: nodes with higher degrees clearly have a bigger chance than nodes with low degrees. These observations complicate the derivation of step-wise delay considerably.

Consequently, in order to be able to derive the rate $\lambda^{(k)}$ and the mean delay of step k ($T_{k,k+1}$), we need to keep track of the (expected) degrees that the infected nodes have at state k. Specifically, we need to derive the following quantities related to spreading over a configuration contact graph: (i) the expected degree of the next node to receive the message at state k, $\mu_d^{new}(k)$; and (ii) the *out degree* at step k, $D^{out}(k)$.

2.3.2.1.1 Mean Degree Assume we are at state k. Let us denote as $\mathbf{p_d}(k)$ the degree distribution of the $N - k$ nodes that *do not* have the packet at state k and $\mu_d(k)$ and $CV_d(k)$

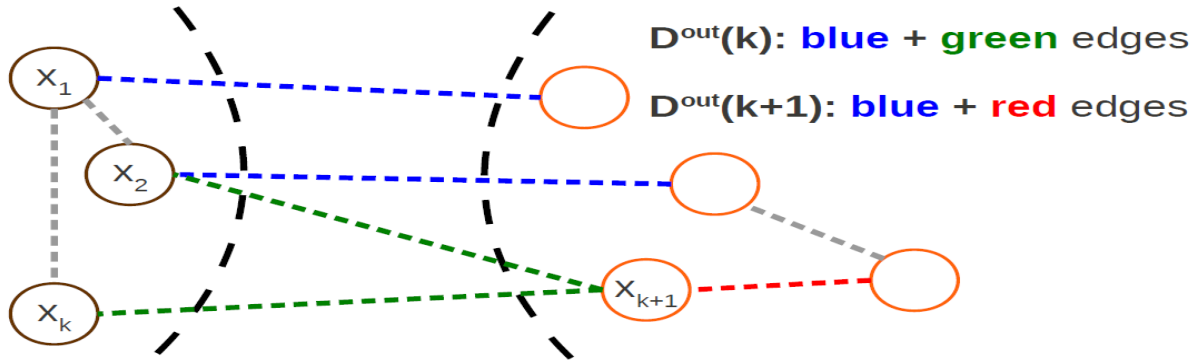


Figure 2.11: Sets of nodes with (left) and without (right) the message at state k. Nodes are represented by circles and edges by the straight lines.

its expectation and coefficient of variation respectively[10]. As we mentioned, not all (uninfected) nodes are equally likely to be the next one infected. As a result, the expected degree of the next infected node is neither equal to μ_d (the original mean degree) nor $\mu_d(k)$.

Result 2. *The expected degree of the next node that will receive the message at step k, is approximately given by*

$$\mu_d^{new}(k) = \mu_d \cdot \left(\frac{N-k-1}{N-1}\right)^{CV_d^2} \cdot (1+CV_d^2) \tag{2.24}$$

Proof. To derive the above result, we need to define and solve an appropriate recursion. Observe that there are $D^{out}(k)$ links across which the infection may proceed from state k to $k+1$ (see Fig. 2.11) and each of these occurs with equal probability (due to equal rates λ). It is a standard result in complex network analysis [102] that the degree distribution of the node reached from that link (i.e. the next node which will receive the message) is[11]:

$$p_d^{new}(k) = \frac{d \cdot p_d(k)}{\sum_d d \cdot p_d(k)} = \frac{d}{\mu_d(k)} \cdot p_d(k) \tag{2.25}$$

Eq. (2.25) implies that the higher degree d a node has, the more probable is that this node will be the next node to receive the message: the probability the new node to have degree d is proportional to $d \cdot p_d(k)$. Now, we can easily derive $\mu_d^{new}(k)$:

$$\mu_d^{new}(k) = \sum_d d \cdot p_d^{new}(k) = \mu_d(k) \cdot \left[1 + CV_d^2(k)\right] \tag{2.26}$$

We can see that the expected degree of the next node infected is higher than the mean degree of all the *uninfected* nodes: $\mu_d^{new}(k) \geq \mu_d(k)$.

To proceed further, we need to know $\mu_d(k)$ and $CV_d^2(k)$ first. To this end, we can set up a recursion for the degree distribution $\mathbf{p_d}(k)$ of the nodes that do not have the message in the

[10]The values of these quantities before the beginning of the spreading, are equal to the values of the initial distribution, i.e. $\mathbf{p}_d(0) = \mathbf{p}_d$, $\mu_d(0) = \mu_d$ and $CV_d(0) = CV_d$.

[11]In the remainder we denote as $p_d(k)$ the probability the degree of a node to be equal to d. Note the difference with the (whole) degree distribution, which is denoted (with a bold symbol) as $\mathbf{p_d}(k)$. I.e., $\mathbf{p_d}(k)$ is a vector : $\mathbf{p_d}(k) = \{p_1(k), p_2(k), \cdots, p_{d_{max}}(k)\}$

next state. Notice that the set of the nodes without the message in state $k+1$ is the same set as in the previous state k, except for the node that just received the message. Hence, we can write for the number of nodes with degree d in states k and $k+1$:

$$[N-(k+1)] \cdot p_d(k+1) = (N-k) \cdot p_d(k) - p_d^{new}(k) \tag{2.27}$$

Substituting in Eq. (2.27) the value of $p_d^{new}(k)$ from Eq. (2.25), we find:

$$p_d(k+1) = \frac{N-k}{N-(k+1)} \cdot p_d(k) - \frac{1}{N-(k+1)} \cdot \frac{d}{\mu_d(k)} \cdot p_d(k) \tag{2.28}$$

In Eq. (2.28), we have expressed $p_d(k+1)$ as a function of $p_d(k)$. Now, it is straightforward to do the same for the expected value, $\mu_d(k+1) = \sum_d d \cdot p_d(k+1)$, and the recursive relation for it, is:

$$\mu_d(k+1) = \mu_d(k) \cdot \left(1 - \frac{CV_d^2(k)}{N-(k+1)}\right) \tag{2.29}$$

where $CV_d^2(k) = \frac{\sigma_d^2(k)}{\mu_d^2(k)} = \frac{\sum_d d^2 \cdot p_d(k) - \mu_d^2(k)}{\mu_d^2(k)}$.

To calculate $\mu_d(k+1)$, the value of $CV_d^2(k)$ is also needed. While we could also set up a recursion to derive the latter, it is proved in Appendix 2.6.5 that it requires knowledge of all higher moments of the degree distribution. To keep things simple and avoid requiring such knowledge (beyond the second moment), we will assume that $CV_d(k) = CV_d \ \forall k$. The conditions for this assumption and its accuracy are discussed in Appendix 2.6.5 and, here, we will only mention the main points which are: (i) the approximation can be accurate for steps k for which it holds $N-k \gg CV_d$, and (ii) it becomes more accurate as the CV_d decreases.

Thus, using $CV_d(k) = CV_d$, and $\mu_d(0) = \mu_d$, Eq. (2.29) gives

$$\mu_d(k) = \mu_d \cdot \prod_{m=0}^{k-1} \left(1 - \frac{CV_d^2}{N-m-1}\right) \tag{2.30}$$

To find an equivalent closed-form expression for Eq. (2.30), we can use the Taylor series approximation for the function $f(x) = e^{-x}$, about $x = 0$, which is $\mathcal{T}(e^{-x}) \approx 1-x$ and is quite accurate for values $0 < x < 0.5$ (with increasing accuracy as x decreases). Then, setting $x = \frac{CV_d^2}{N-m-1}$ (the accuracy condition is satisfied for the states k for which $N-k > 2 \cdot CV_d^2$ and thus more accuracy can be achieved for lower values of CV_d), we can write for Eq. (2.30)

$$\begin{aligned}
\mu_d(k) \approx \mu_d \cdot \prod_{m=0}^{k-1} e^{-\frac{CV_d^2}{N-m-1}} &= \mu_d \cdot exp\left\{-CV_d^2 \cdot \sum_{m=0}^{k-1} \frac{1}{N-m-1}\right\} \\
&= \mu_d \cdot exp\left\{-CV_d^2 \cdot \sum_{\ell=N-k}^{N-1} \frac{1}{\ell}\right\} \approx \mu_d \cdot exp\left\{-CV_d^2 \cdot [\ln(N-1) - \ln(N-k-1)]\right\} \\
&= \mu_d \cdot exp\left\{\ln\left[\left(\frac{N-k-1}{N-1}\right)^{CV_d^2}\right]\right\} = \mu_d \cdot \left(\frac{N-k-1}{N-1}\right)^{CV_d^2}
\end{aligned} \tag{2.31}$$

where we have used the *harmonic series* approximation[12], which holds for $N-k \gg 1$ and whose accuracy increases for larger values of $N-k$.

Substituting Eq. (2.31) in Eq. (2.26) gives us Result 2. □

[12] $\sum_{n=1}^{k} \frac{1}{n} \approx \ln(k) + \gamma$, where γ is the Euler-Mascheroni constant.

2.3.2.1.2 Out Degree

Result 3. *The mean value of the out degree at step k, $D^{out}(k)$, is approximately given by*

$$\overline{D}^{out}(k) = (N-k)\mu_d \left[\left(\frac{N-k}{N-1}\right)^{CV_d^2} - \left(\frac{N-2}{N-1}\right)\left(\frac{N-k}{N-1}\right)^{2CV_d^2+1} \right] \tag{2.32}$$

To derive Result 3 we have followed a similar method as before to form a recursion:

$$\overline{D}^{out}(k+1) = \overline{D}^{out}(k) + [\mu_d^{new}(k) - 2] - 2\left[\overline{D}^{out}(k) - 1\right] \frac{\mu_d^{new}(k) - 1}{(N-k)\cdot\mu_d(k) - 1} \tag{2.33}$$

The details about the setup and solution of Eq. (2.33) can be found in Appendix 2.6.6. We will only provide here an intuitive sketch of proof based on a simple example.

In Fig. 2.11, the set of nodes with the message is $\mathbf{C}(k) = \{x_1, ..., x_k\}$ and the out degree of step k is given by the number of edges that connect the nodes $\in \mathbf{C}(k)$ with the nodes $\notin \mathbf{C}(k)$ (blue+green edges). If we denote as x_{k+1} the next node to receive the message and assume that the node x_2 disseminates the message to x_{k+1}, the out degree of the next step, $D^{out}(k+1)$, is calculated as following:

From the value of $D^{out}(k)$ we have to subtract the number of edges that connect the nodes $\in \mathbf{C}(k)$ with the node x_{k+1} (green edges). Let us denote this number as N_1. Then we have to add the number of the edges of the new node x_{k+1} that connect it with the nodes $\notin \mathbf{C}(k)$ (red edges) and we denote this number as N_2. It is evident that $N_2 = d^{new} - N_1$, where d^{new} is the degree of the node x_{k+1}. So we can write:

$$D^{out}(k+1) = D^{out}(k) - N_1 + N_2 = D^{out}(k) + d^{new} - 2\cdot N_1 \tag{2.34}$$

To estimate the number of the edges that connect the nodes $\in \mathbf{C}(k)$ with the node x_{k+1} (green edges), i.e. N_1, we should consider that each of the edges of $D^{out}(k)$, except for the one that connected to x_{k+1}, is connected with another edge of x_{k+1} with probability $\frac{d^{new}(k)-1}{(N-k)\cdot\mu_d(k)-1}$, where $d^{new}(k) - 1$ is the number of the unoccupied edges of x_{k+1} and $(N-k)\cdot\mu_d(k) - 1$ is the total number of edges of the nodes $\notin \mathbf{C}(k)$. We do not take into account the probability of double edges or self-loops, because this probability for large networks is almost zero [102]. So the expectation of N_1 will be

$$E[N_1] = 1 + (D^{out}(k) - 1)\cdot\frac{d^{new}(k) - 1}{(N-k)\cdot\mu_d(k) - 1} \tag{2.35}$$

Now, from equations Eq. (2.34) and Eq. (2.35), we can prove Eq. (2.33)[13]. Furthermore, using Eq. (2.26) (with $CV_d(k) \approx CV_d$) and assuming that the minimum degree, d_{min}, of the network is much larger than 1, which also implies that $\mu_d^{new}(k), D^{out}(k) \geq d_{min} \gg 1$, we can write for Eq. (2.33):

$$\overline{D}^{out}(k+1) = \overline{D}^{out}(k)\cdot\left[1 - 2\frac{1+CV_d^2}{N-k}\right] + (1+CV_d^2)\cdot\mu_d(k) \tag{2.36}$$

[13]Note the difference in notation between $D^{out}(k)$ and its mean value $\overline{D}^{out}(k)$.

The solution of Eq. (2.36), for $\overline{D}^{out}(1) = \mu_d$, is the Result 3.

Piecewise Formula: The above result provides us with a closed form expression for the *mean* value of the out degree $D^{out}(k)$, at step k, which allows us to calculate the necessary transition rates $\lambda^{(k)}$ in Eq.(2.23). However, it is based on Eq. (2.31) that was derived using some assumptions ($N - k \gg 1$ and $CV_d(k) = CV_d$), under which we tend to underestimate $\mu_d(k)$. Specifically, for some distributions p_d, Eq. (2.31) might produce, in the last steps of the recursion, unacceptably small values for $\mu_d(k)$. We can easily correct this by explicitly forcing $\mu_d(k) \geq d_{min}$ (which always holds). Then, it can be proved (Appendix 2.6.7) that a better approximation for $D^{out}(k)$ is given by the following piecewise result:

Result 4. *The mean value of the out degree is calculated by Result 3 for $k \leq k_{stop}$, and by*

$$\overline{D}^{out}(k) = (N-k)^2 \cdot \left[\frac{D_{stop} - d_{min} \cdot (N - k_{stop})}{(N - k_{stop})^2} + \frac{d_{min}}{N-k} \right] \tag{2.37}$$

for $k > k_{stop}$, where $k_{stop} = \left[1 - \left(\frac{d_{min}}{\mu_d} \right)^{\frac{1}{CV_d^2}} \right] \cdot (N-1)$, and D_{stop} is computed by setting $k = k_{stop}$ in the expression of Result 3.

2.3.2.1.3 Spreading Delay To conclude our derivation, let us look back at our initial equation for the rates of Fig.2.10, $\lambda^{(k)} = \lambda \cdot D^{out}(k)$. Note that we have derived thus far the *expected* value for $D^{out}(k)$. Yet, $D^{out}(k)$ is a random variable depending on $\mathbf{C}(k)$, the actual set of the k nodes that have the message at state k. *Given* $\mathbf{C}(k)$, the delay of step k, $T_{k,k+1}$, is an exponential random variable with rate $\lambda^{(k)} = \lambda \cdot D^{out}(k)$. Thus,

$$E\left[T_{k,k+1} | \mathbf{C}(k)\right] = \tfrac{1}{\lambda \cdot D^{out}(k)}, \tag{2.38}$$

and using the properties of conditional expectation, we get the expected delay of step k:

$$E\left[T_{k,k+1}\right] = \sum_{\mathbf{C}(k)} \frac{1}{\lambda \cdot D^{out}(k)} \cdot P\{\mathbf{C}(k)\} = \frac{1}{\lambda} \cdot E\left[\frac{1}{D^{out}(k)} \right] \tag{2.39}$$

We cannot, in general, replace $E\left[\frac{1}{D^{out}(k)}\right]$ above, which is hard to calculate, with $\frac{1}{\overline{D}^{out}(k)}$, which follows directly from Eq.(2.32) and (2.37). In fact, Jensen's inequality suggests that $\frac{1}{\overline{D}^{out}(k)} \leq E\left[\frac{1}{D^{out}(k)}\right]$.

To proceed with our approximation, we resort to the *Delta method* [103] for approximating the expectation of functions of random variables. Here, the random variable is $X = D^{out}(k)$ and we need to compute (Eq. (2.39)) the expectation of the function $f(X) = \frac{1}{X}$. We can approximate $f(X)$ with a Taylor series expansion about the mean value $E[X] = \overline{D}^{out}(k)$. Finally, by keeping only the first few terms of this series and taking their expectation, we can more easily express $E[T_{k,k+1}]$ as a function of moments of $D^{out}(k)$. Specifically, considering the first two terms of the expansion, we get

$$E\left[T_{k,k+1}\right] = \frac{1}{\lambda} \cdot E\left[\frac{1}{D^{out}(k)} \right] \approx \frac{1}{\lambda \cdot \overline{D}^{out}(k)} \tag{2.40}$$

Now, in Eq. (2.40), we can calculate the expected *step delay* by substituting the value of Result 3 or Result 4.

The accuracy of the Delta method and the above approximation is higher, if the mass of the random variable $X = D^{out}(k)$ is concentrated around its mean $\overline{D}^{out}(k)$ [103]. It is known that, in a configuration model network, the network structural properties and the properties of processes on the network becoming concentrated more and more narrowly around their mean value [102], as the network size increases. Therefore, the larger the network size N, the higher the accuracy of the approximation. Furthermore, if increased accuracy is desired, more terms in the Taylor series above could be used (by deriving a few higher moments of $D^{out}(k)$).

2.3.2.2 Model Validation

In order to validate our model, we compare the theoretical results we derived, against a sample of simulations for both synthetic and real-world networks.
Synthetic Simulations: At first, we created various synthetic scenarios conforming to our model (Def. 6). For each scenario, the procedure we follow, is:

1. We choose an initial degree distribution $\mathbf{p_d}$.
2. With the configuration model we create 50 different networks (contact graphs) and for each pair of nodes in a network we create a sequence of contact events with inter-contact times drawn from an exponential distribution with rate $\lambda = 1$.
3. For each network, we generate 1000 messages at random times and at random source nodes and start the spreading.
4. We calculate the average values, over all networks and spreading processes of the specific scenario, of the *out degree*, $\overline{D}^{out}(k)$, and *step delay*, $E[T_{k,k+1}]$, of each step.

To choose realistic parameters for the degree distributions in our scenarios, we analysed contact graphs of real-world mobile social networks[14] and found that the degrees follow either a uniform or right-skewed distribution with CV_d in the range $[0.6, 0.85]$ (details for the scenarios are given in Table 2.4).

Table 2.4: Parameters of the contact graphs of four real-world scenarios.

TRACE	network size N	μ_d	CV_d
Sigcomm 2009	76	25.5	0.6
SocioPatterns	111	7.6	0.85
Cabspotting	536	120	0.74
Infocom 2006	98	32	0.61

In Fig. 2.12 we present the *out degree* for each step in two scenarios with 1000 nodes. We compare the simulation values with the theoretical (Results 3 and 4). We can see that the achieved accuracy is significant. As expected, in the scenario with higher CV_d the accuracy is

[14]The traces are available at:
1) *Sigcomm 2009*: http://crawdad.cs.dartmouth.edu/thlab/sigcomm2009
2) *SocioPatterns*: http://www.sociopatterns.org/
3) *Cabspotting*: http://crawdad.cs.dartmouth.edu/epfl/mobility
4) *Infocom 2006*: http://crawdad.cs.dartmouth.edu/cambridge/haggle

lower, especially for the last steps, because the approximations we did in the derivation of the theoretical results are less accurate as the CV_d increases. Also, in Fig. 2.12(a) there was not need to use the piecewise formula (Result 4) and in the second case, Fig. 2.12(b), it should be used only for the last 25% of the steps. The corresponding values for $D^{out}(k)$ that a *fully-meshed* network model would predict are very far from the simulated values (e.g. for the 500^{th} step it gives a value 15 times larger). We therefore also compare our results to a baseline model: a *regular graph* with the same number of edges as our network, but where every node has the same degree. Fig. 2.12 confirms that our model performs significantly better.

In Table 2.5 we present the *average relative errors* for $D^{out}(k)$, defined as

$$E\left[\frac{|D^{out}(k)_{sim} - D^{out}(k)_{th}|}{D^{out}(k)_{sim}}\right]$$

for four networks (of which the two correspond to the results presented in Fig. 2.12) of 1000 nodes and similar μ_d values. We show the average relative error for the first 250, 500 and 750 steps and the total (over all steps). The more steps we consider, the higher the error is. This comes of the fact that our theoretical results are less accurate for the last steps of the spreading. It can be seen that for networks with lower CV_d the error is lower. For example, for $CV_d = 0.31$, the error is insignificant, even for the last steps. For the extreme case of $CV_d = 1.29$ [15] the error is not negligible. However, our prediction is still acceptable, if we consider the heterogeneity this scenario has.

Table 2.5: Relative step error of $D^{out}(k)$ on different network scenarios.

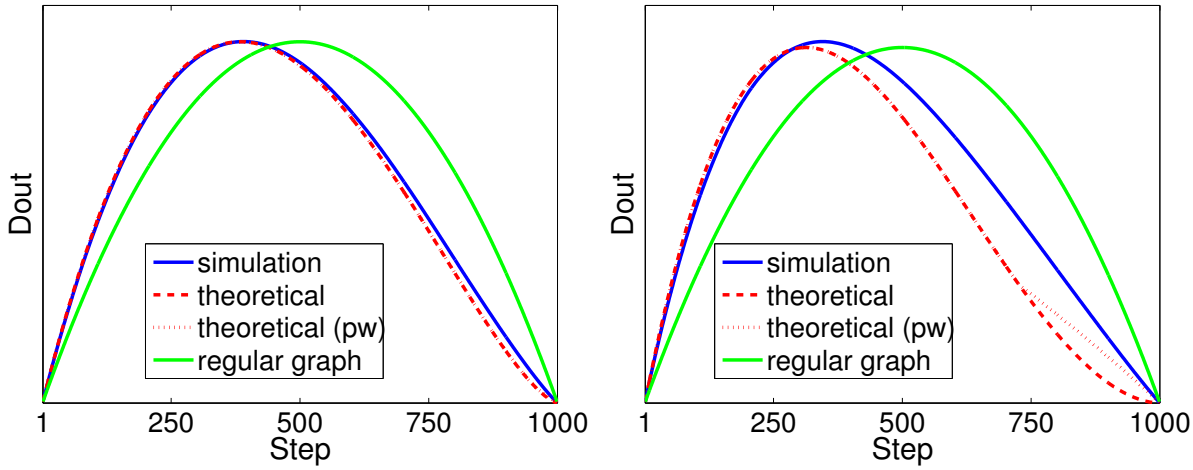
	250 steps	500 steps	750 steps	over all steps
$CV_d = 0.31$	1%	2%	2%	2%
$CV_d = 0.65$	1%	1%	2%	6%
$CV_d = 0.92$	4%	4%	11%	15%
$CV_d = 1.29$	14%	18%	27%	29%

Fig. 2.13 shows the *aggregate step delay* (i.e. the time the message needs to be spread in k nodes) for two synthetic scenarios: (a) network with with 100 nodes, $\mu_d = 23$ and $CV_d = 0.71$; and (b) network with 500 nodes, $\mu_d = 30$ and $CV_d = 1.16$ [16]. Similarly to the results for $D^{out}(k)$, it can be seen also here that the theoretical *aggregate step delay* is close to the simulated value for almost every step.

Synthetic Simulations - Heterogeneous Rates: Further, we investigate the performance of our model in networks with heterogeneous contact rates (different λ_{ij} for each pair). We create synthetic scenarios and run simulations as before. The only difference is the generation of the contact events, where, now, the inter-contact times are exponentially distributed but with a different rate for each pair. We chose λ_{ij} to follow a log-normal distribution with $\mu_\lambda = 1$ and $\sigma_\lambda^2 = 3$.

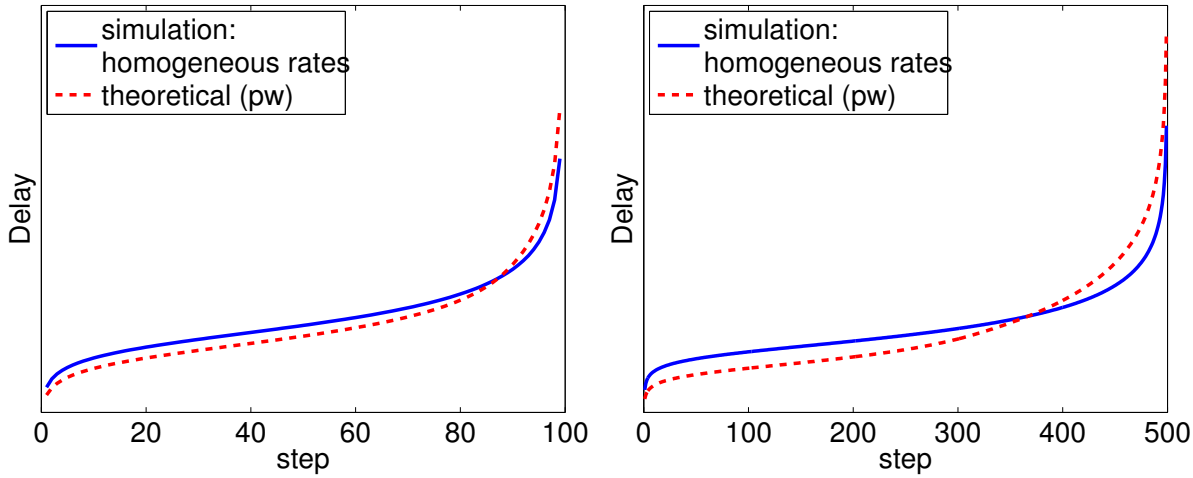
[15] We characterise it as an extreme case, as the min and max degrees in this network are 22 and 968, respectively, in order to have a CV_d value as high as possible.

[16] It is the higher variance we could achieve among all the scenarios of 100 and 500 nodes, respectively. The degree distribution was highly skewed and the maximum degree in the network was almost equal to the network size, $d_{max} = 100$ and $d_{max} = 500$ for the two cases.



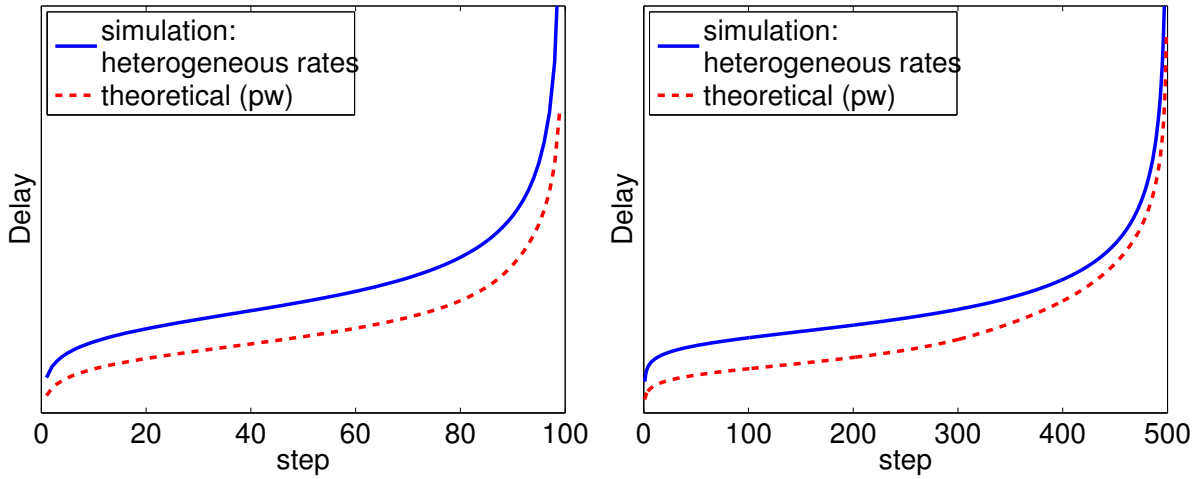
(a) $CV_d = 0.65$ (b) $CV_d = 0.92$

Figure 2.12: $D^{out}(k)$ of each step in two scenarios with 1000 nodes.



(a) $\mu_d = 23$, $CV_d = 0.71$ (b) $\mu_d = 30$, $CV_d = 1.16$

Figure 2.13: Aggregate step delay. Synthetic simulations in scenarios with: (a) 100 nodes, $\mu_d = 23$ and $CV_d = 0.71$; and (b) network with 500 nodes, $\mu_d = 30$ and $CV_d = 1.16$.



(a) $\mu_d = 23$, $CV_d = 0.71$ (b) $\mu_d = 30$, $CV_d = 1.16$

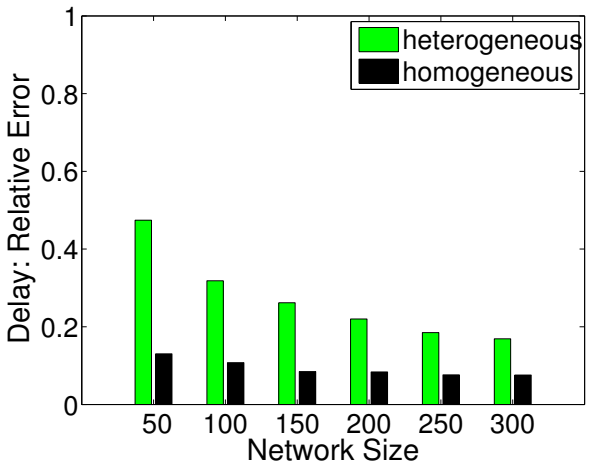
Figure 2.14: Aggregate step delay. Synthetic simulations in scenarios with heterogeneous contact rates: (a) 100 nodes, $\mu_d = 23$ and $CV_d = 0.71$; and (b) network with 500 nodes, $\mu_d = 30$ and $CV_d = 1.16$.

The results for the *aggregate step delay* are presented in Fig. 2.14. The scenarios presented are the corresponding to the homogeneous-rates scenarios of Fiq. 2.13. As can be seen in Fig. 2.14, simulation and theoretical results diverge more for the heterogeneous contact rate scenario.

This divergence is more clearly seen in Fig. 2.15, which shows the relative error of the average *aggregate step delay* over all the steps, i.e. $E\left[\frac{|D_{sim}-D_{th}|}{D_{sim}}\right]$ where D denotes the aggregate step delay. We present six scenarios of different network sizes. For each scenario we chose a bounded pareto degree distribution with minimum value $d_{min} = 0.1 \cdot N$ (N is the network size), $d_{max} = N$ and shape factor the one that resulted in the higher CV_d. These represent the worst case parameters (among the ones we observed in real traces) that most hurt the accuracy of our model. Nevertheless, in the homogeneous scenarios, the error is very low (below 10% for almost all the networks) and, in the heterogeneous scenarios, it is always higher, but decreases for larger network sizes. For a network with 300 nodes, it becomes approximately 20%, which is rather satisfying, given the high variability in both the degrees and rates in this scenarios.

Real-world Networks: After evaluating the accuracy of our model in a range of different (regarding the network size, degree distribution, contact rates) synthetic scenarios, we present here the results of simulations on real-world traces. It is of interest to see to what extent our model can capture the quantities of interest in a real-world scenario, where the assumptions do not hold exactly, as we have noted community structure (i.e. the *clustering coefficient* [102] is $27 - 50\%$ more than in the corresponding configuration model network), heterogeneous contact rates and non-Poisson contact events (e.g. less contacts during night hours).

Fig. 2.16 shows the results of 1000 simulation runs on the mobility trace from the 4 days iMotes experiment during Infocom 2006 [125], which contains traces of Bluetooth sightings of 78 mobile and 20 static nodes. In Fig. 2.16(a) it can be seen that the theoretically predicted *out degree* only differ slightly, except for some last steps, from the simulation's average. Thus we can infer, that despite the community structure of this network, our model can still capture the way the spreading proceeds among nodes with different degree. Fig. 2.16(b) shows the *aggregate step delay*. We can see that the accuracy is good for more than half of the steps. However, in



(a) Delay: Relative Errors

Figure 2.15: Relative errors of the delay averaged over all the steps in scenarios with Homogeneous and Heterogeneous contact rates for 6 different network sizes.

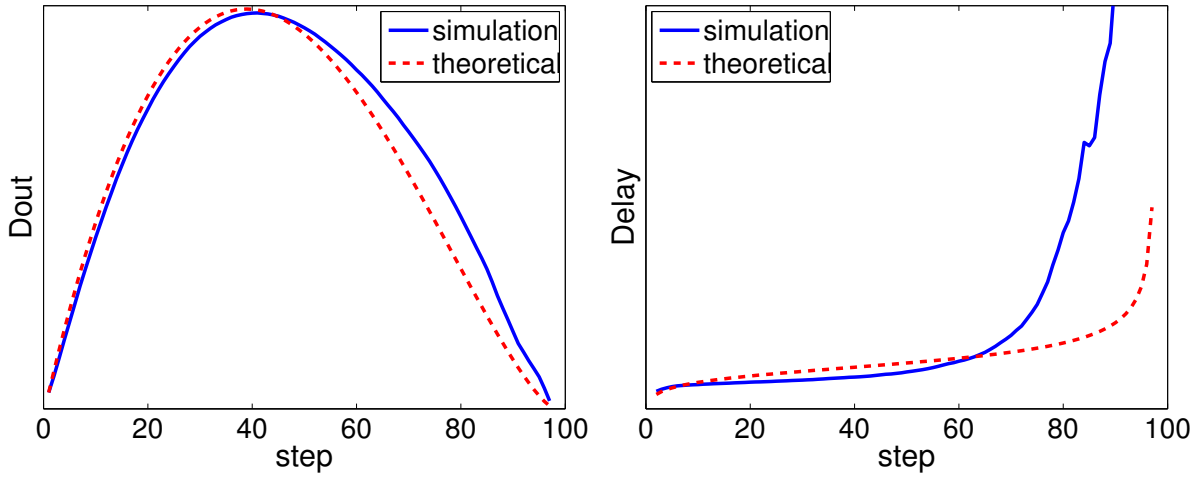
the following steps our theoretical results are far from the observed delay. An explanation for this, is the correlation between the contact events of different pairs which affects the spreading process (e.g. in conference events there are much more contact events than during night hours).

We have observed similar good accuracy for the first 70-75% steps and divergence subsequently, in other traces as well. In Fig. 2.17 we present the results of 1000 simulation runs on the mobility trace *Cabspotting* [118], which contains GPS coordinates from 536 taxi cabs collected over 30 days in San Francisco.

2.4 Related Work

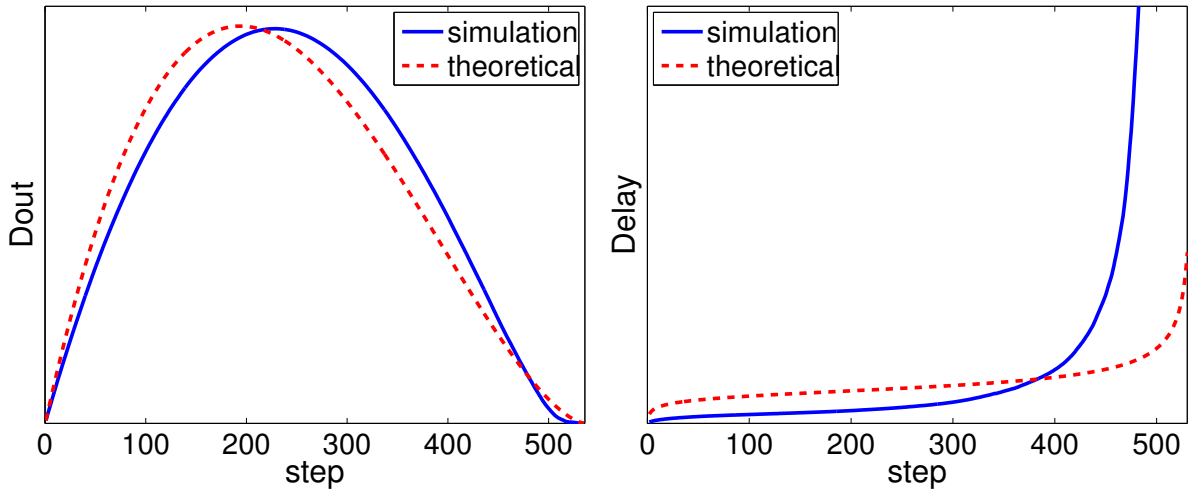
Models for epidemic spreading of diseases [102] and/or computer malware [140], were early derived, based on the well known SIR model, and studied widely. In DTNs, efforts to analyze the performance of epidemic routing and other protocols also abound. Stochastic analyses, like the one in [43], define a Markov chain as in Fig. 2.1, in order to give closed form results for epidemic and 2-hop routing. Fluid models [49, 72, 143], take an approach similar to the SIR model in biology, and define the number of messages in the network as a continuous function (of time). Then, ordinary differential equations (ODEs) are used to derive expressions for the total delay, delivery probability etc. While these models provide closed form results and thus can be used in tuning protocol parameters (e.g. gossiping probability [143], number of copies [129], TTL [89], they all assume a homogeneous network with a common meeting rate for every pair of nodes.

Recent studies on real network traces [25, 36, 56] suggest that the homogeneity assumption is not true. To overcome this limitation, a number of works introduced heterogeneity in contact network models, by allowing different meeting rates for each node pair [36, 60, 76, 113, 132]. Yet, most of these works use the heterogeneous model to design new, better protocols (e.g. multicast [36] or unicast [60]) that take heterogeneity into account, but do not analyze performance. One exception is [76], but only for the cases of direct transmission and 2-hop routing. To our



(a) $D^{out}(k)$ (b) Delay

Figure 2.16: Simulations on Infocom 2006 trace: 96 nodes, $\mu_d = 33$, $CV_d = 0.6$



(a) $D^{out}(k)$ (b) Delay

Figure 2.17: Simulations on Cabspotting trace: 536 nodes, $\mu_d = 120$, $CV_d = 0.74$

best knowledge, the work closer to this paper is that of [113], where a very generic contact graph is considered. However, due to the large generality of the contact model, only upper bounds for the delay can be provided.

In our work, while we allow arbitrary link rates between nodes, similarly to [36, 76, 113], we restrict the underlying contact graph model, in order to derive closed-form analytical expressions. We validated our results with synthetic simulations for the targeted contact classes, as previous work did [43, 129, 143], but also demonstrate their applicability in real networks.

As a final note, numerous studies exist in the field of Complex Networks (including theoretical biology, epidemiology, etc.) on epidemic processes over various complex network models (e.g. [7, 71, 96, 100, 101, 140]). However, the majority of these works focus on deriving thresholds above which the epidemic will spread and their results usually consider infinite time. Additionally, it is not always feasible to apply them in real scenarios for predicting the spreading delay, as they require, e.g., the complete knowledge of the underlying contact graph [140] or the exact degree distribution [96,100,101], information that is usually very difficult, if not impossible, to estimate in real-time MSNs. On the contrary, our results include only average statistics of the network, e.g. moments of the contact rates distribution (μ_λ, CV_λ) or the degree distribution (μ_d, CV_d), which are easier to calculate, estimate or infer.

2.5 Conclusions

In this chapter, we have considered a generic class of heterogeneous contact models, and have derived both asymptotic results and simple closed form approximations for epidemic spreading. We also extended our analysis for networks with sparse graphs and graphs with arbitrary degree distributions, where neighbors contact randomly. From the validation of the model against synthetic simulations and real traces we can conclude that: (a) simple delay expressions that can be used for performance prediction and protocol optimization (and require only partial knowledge of the network characteristics), exist not only for the homogeneous contact case; (b) performance predictions that are accurate qualitatively, and (somewhat more surprisingly) sometimes quantitatively also, can be made even for a number of real scenarios, despite the highly more complex structure of the latter.

A further utility of the Heterogeneous Contact Network model is that it makes possible to take into account additional properties, related to (the heterogeneous) social characteristics of users in MSNs. To this direction, in the following chapter, based on this model and the theoretical results we derived, we investigate the effects of another social dimension of MSNs, namely the users' *social cooperation* or *social selfishness*.

2.6 Appendix: Supplementary Theoretical Results and Proofs

2.6.1 Proof of Lemma 1

Proof. The sum of the contact rates between infected nodes ($i \in \mathbf{C}_k^m$) and susceptible nodes ($j \notin \mathbf{C}_k^m$) at step k, is related to the respective sum of the previous step as following

$$S_k^m = S_{k-1}^m - \sum_{i \in \mathbf{C}_{k-1}^m, j=n_k} \lambda_{ij} + \sum_{i \notin \mathbf{C}_k^m, j=n_k} \lambda_{ij} \tag{2.41}$$

where:

(i) We denote as n_k the k^{th} infected node (i.e. the node infected at the transition between step $k-1$ and step k).

(ii) $\sum_{i\in\mathbf{C}_{k-1}^m, j=n_k}\lambda_{ij}$ is the sum of the contact rates between the infected nodes at step $k-1$ and node n_k. These rates are included in the sum S_{k-1}^m, but are not included to the sum S_k^m (since at step k they belong to the set of contact rates between infected nodes), and, hence, we subtract them in Eq. (2.41).

(iii) $\sum_{i\notin\mathbf{C}_k^m, j=n_k}\lambda_{ij}$ is the sum of the contact rates between node n_k and the susceptible nodes at step k. These rates are included in the sum S_k^m, but are not included to the sum S_{k-1}^m (since at step $k-1$ they belong to the set of contact rates between susceptible nodes), and, hence, we add them in Eq. (2.41).

Now, we first split the sum $\sum_{i\in\mathbf{C}_{k-1}^m, j=n_k}\lambda_{ij}$ in two terms

$$\sum_{i\in\mathbf{C}_{k-1}^m, j=n_k}\lambda_{ij} = \lambda_{k-1}^{next} + \sum_{i\in\mathbf{C}_{k-1}^m, j=n_k, \lambda_{ij}\neq\lambda_{k-1}^{next}}\lambda_{ij} = \lambda_{k-1}^{next} + S_{k-1}^{next} \tag{2.42}$$

where we denoted as λ_{k-1}^{next} the meeting rate between the next node to get the message (i.e. n_k) and the node who infected him, and

$$S_{k-1}^{next} = \sum_{i\in\mathbf{C}_{k-1}^m, j=n_k, \lambda_{ij}\neq\lambda_{k-1}^{next}}\lambda_{ij} \quad \text{(comprising } k-2 \text{ terms)} \tag{2.43}$$

We further denote

$$S_{next}^k = \sum_{i\notin\mathbf{C}_k^m, j=n_k}\lambda_{ij} \quad \text{(comprising } N-k \text{ terms)} \tag{2.44}$$

Using Eq. (2.42), Eq. (2.43) and Eq. (2.44), we can write Eq. (2.41) as

$$S_k^m = S_{k-1}^m - \lambda_{k-1}^{next} - S_{k-1}^{next} + S_{next}^k \tag{2.45}$$

Based on the above recursive relation, in the remainder, we calculate the expectation and variance of S_k. Before proceeding, let us first define the following quantities (for $k = 1, \cdots, N-1$)

$$\mu_k = \frac{E[S_k]}{k(N-k)} \tag{2.46}$$

$$\sigma_k^2 = \frac{Var\,[S_k]}{k(N-k)} \tag{2.47}$$

Expectation

Taking the expectation in Eq. (2.45), gives[17]

$$E[S_k] = E[S_{k-1}] - E[\lambda_{k-1}^{next}] - E[S_{k-1}^{next}] + E[S_{next}^k] \tag{2.48}$$

Now, we express the terms in the right side of Eq. (2.48), as following:

[17]In Eq. (2.42)-Eq. (2.45), the quantities λ_{k-1}^{next}, S_{k-1}^{next} and S_{next}^k correspond to the sets $\mathbf{C}_{k-1}^m$ and $\mathbf{C}_k^m$. We dropped the superscripts m to avoid notation complexity. In the remainder, the expectations are taken over all the possible values (for different m) that these quantities can take.

(i) At first, by the definition of Eq. (2.46), we can write

$$E[S_{k-1}] = (k-1)(N-k+1)\cdot\mu_{k-1} \tag{2.49}$$

(ii) The probability the node pair $\{x,y\}$, $x\in\mathbf{C}_{k-1}^m, y\notin\mathbf{C}_{k-1}^m$ (among all the node pairs $\{i,j\}$, $i\in\mathbf{C}_{k-1}^m, j\notin\mathbf{C}_{k-1}^m$) to be the pair through which the message is spread at step $k-1$ (i.e. y is the k^{th} node that is infected, and it is infected by node x) is proportional to its contact rate λ_{xy} (because intercontact intervals are exponentially distributed). Hence, we can, equivalently, write

$$P\{\lambda_{k-1}^{next}=\lambda_{xy}|x\in\mathbf{C}_{k-1}^m, y\notin\mathbf{C}_{k-1}^m\} = \frac{\lambda_{xy}}{\sum_{i\in\mathbf{C}_{k-1}^m}\sum_{j\notin\mathbf{C}_{k-1}^m}\lambda_{ij}} = \frac{\lambda_{xy}}{S_{k-1}^m} \tag{2.50}$$

From Eq. (2.50), it is easy to see that the rate λ_{k-1}^{next} will be on average larger than the average rate between node pair $\{i,j\}$, $i\in\mathbf{C}_{k-1}^m, j\notin\mathbf{C}_{k-1}^m$, i.e.

$$E[\lambda_{k-1}^{next}] \geq \mu_{k-1} \tag{2.51}$$

Combining the previous inequality with the fact that the contact rates take values in the interval $[\lambda_{min},\lambda_{max}]$, we can write

$$E[\lambda_{k-1}^{next}] = \mu_{k-1} + \epsilon_{k-1}^{next} \tag{2.52}$$

where

$$0\leq\epsilon_{k-1}^{next}\leq\lambda_{max}-\mu_{k-1} \quad \epsilon_{k-1}^{next} = O\left(\lambda_{max}\right) \tag{2.53}$$

(iii) Since S_{k-1}^{next} is a sum of $k-2$ independent random variables with mean value μ_{k-1}, it follows that

$$E[S_{k-1}^{next}] = (k-2)\cdot\mu_{k-1} \tag{2.54}$$

Remark: In fact the mean value of the terms in the sum S_{k-1}^{next} is slightly different (smaller) than μ_{k-1}, because the rate λ_{k-1}^{next} is not taken into account. Thus, the exact expression for Eq. (2.54) is $E[S_{k-1}^{next}] = (k-2)\cdot\mu_{k-1}+\epsilon_{k-1}^{*}$. However, it can be shown that

$$\epsilon_{k-1}^{*} \leq \frac{(\lambda_{max}-\mu_{k-1})\cdot(k-1)}{(k-1)(N-k+1)-1} \quad\Rightarrow\quad \epsilon_{k-1}^{*} = O\left(\frac{\lambda_{max}}{N-k+1}\right) \tag{2.55}$$

In Eq. (2.55) it is easy to see that $\epsilon_{k-1}^{*}\ll(k-2)\cdot\mu_{k-1},\forall k$ (for large N). Therefore, in the remainder, we ignore it[18].

(iv) It holds for each step k that the rates $\lambda_k^{out}\in\{\lambda_{ij}: i\notin\mathbf{C}_k^m, j\notin\mathbf{C}_k^m\}$ are independent of the spreading process. Thus, they are distributed with the initial contact rates distribution $f_\lambda(\lambda)$, which means that

$$E[\lambda_k^{out}] = E[\lambda] = \mu_\lambda \tag{2.56}$$

$$Var\left[\lambda_k^{out}\right] = Var\left[\lambda\right] = \sigma_\lambda^2 \tag{2.57}$$

Therefore, from Eq. (2.56) it follows that the expectation of the sum S_{next}^k, which consists of $N-k$ contact rates between nodes that are not infected in step $k-1$ (i.e. $\notin\mathbf{C}_{k-1}^m$), is equal to

$$E[S_{next}^k] = (N-k)\cdot E[\lambda_{k-1}^{out}] = (N-k)\cdot\mu_\lambda \tag{2.58}$$

[18] Additionally, since in Eq. (2.48) we consider $\epsilon_{k-1}^{next} = O\left(\lambda_{max}\right)$ (Eq. (2.53)), we can omit ϵ_{k-1}^{*}, for which it holds $\epsilon_{k-1}^{*} = O\left(\frac{\lambda_{max}}{N-k+1}\right)\leq\epsilon_{k-1}^{next}$.

Substituting in Eq. (2.48) the expressions we derived in (i)-(iv) (Eq. (2.49), Eq. (2.52), Eq. (2.54) and Eq. (2.58)), we get

$$\begin{aligned} E[S_k] &= (k-1)(N-k+1)\mu_{k-1} - (\mu_{k-1} + \epsilon_{k-1}^{next}) - (k-2)\mu_{k-1} + (N-k)\mu_\lambda \\ &= (k-1)(N-k+1)\cdot\mu_{k-1} - (k-1)\cdot\mu_{k-1} - \epsilon_{k-1}^{next} + (N-k)\cdot\mu_\lambda \\ &= (k-1)(N-k)\cdot\mu_{k-1} - \epsilon_{k-1}^{next} + (N-k)\cdot\mu_\lambda \\ &= k(N-k)\cdot\left(\frac{(k-1)\cdot\mu_{k-1}+\mu_\lambda}{k} - \frac{\epsilon_{k-1}^{next}}{k(N-k)}\right) \end{aligned} \tag{2.59}$$

or

$$E[S_k] = k(N-k)\cdot\left(\frac{(k-1)\cdot\mu_{k-1}+\mu_\lambda}{k} - \epsilon_k^{'}\right) \tag{2.60}$$

where

$$\epsilon_k^{'} = O\left(\frac{\lambda_{max}}{k(N-k)}\right) \tag{2.61}$$

Now, to calculate $E[S_k]$ for every step k, we start from the first step ($k=1$), where S_1^m is a sum of $N-1$ i.i.d. random variables λ_{ij} with mean value μ_λ (by the definition of the mobility class). Therefore,

$$E[S_1] = (N-1)\cdot\mu_\lambda \tag{2.62}$$

For the second step ($k=2$), substituting Eq. (2.62) in Eq. (2.60), gives

$$E[S_2] = 2(N-2)\cdot\mu_\lambda\cdot(1-\epsilon_2^{'}), \quad \epsilon_2^{'} = O\left(\frac{\lambda_{max}}{N-1}\right) \tag{2.63}$$

Finally, following the same process recursively, for $k=3,4,...$, it is easy to show that $\forall k$ it holds

$$E[S_k] = k(N-k)\cdot\mu_\lambda\cdot(1-\epsilon_k), \quad \epsilon_k = O\left(\frac{\lambda_{max}}{N-1}\right) \tag{2.64}$$

which proves the first part of Lemma 1.

Variance

Taking the variances in Eq. (2.45), gives [121][19]

$$\begin{aligned} Var[S_k] = Var[S_{k-1}] + Var\left[\lambda_{k-1}^{next}\right] + Var\left[S_{k-1}^{next}\right] + Var\left[S_{next}^{k}\right] \\ - 2\cdot Cov\left[S_{k-1}, \lambda_{k-1}^{next}\right] - 2\cdot Cov\left[S_{k-1}, S_{k-1}^{next}\right] \end{aligned} \tag{2.65}$$

We proceed similarly to the derivation of the expectation, and express the terms in the right side of Eq. (2.65), as following:
(i) At first, by the definition of Eq. (2.47), we can write

$$Var[S_{k-1}] = (k-1)(N-k+1)\cdot\sigma_{k-1}^2 \tag{2.66}$$

[19]The covariances of independent variables are zero, and, thus, we do not include them in Eq. (2.65).

(ii) Similarly to Eq. (2.54)-Eq. (2.55), it follows that

$$Var\left[S_{k-1}^{next}\right] = (k-2)\cdot\sigma_{k-1}^2 + \delta_{k-1}^* \tag{2.67}$$

and because δ_{k-1}^* is small, we ignore it:

$$Var\left[S_{k-1}^{next}\right] = (k-2)\cdot\sigma_{k-1}^2 \tag{2.68}$$

(iii) Making similar arguments as in Eq. (2.58) and using Eq. (2.57), it follows that

$$Var\left[S_{next}^{k}\right] = (N-k)\cdot\sigma_{\lambda}^2 \tag{2.69}$$

(iv) The covariance of two random variables is given by the expression

$$Cov\left[X,Y\right] = E[X\cdot Y] - E[X]\cdot E[Y]$$

Therefore, for the first covariance appearing in the sum of Eq. (2.65) we can write

$$Cov\left[S_{k-1},\lambda_{k-1}^{next}\right] = E[S_{k-1}\cdot\lambda_{k-1}^{next}] - E[S_{k-1}]\cdot E[\lambda_{k-1}^{next}] \tag{2.70}$$

Since S_{k-1} is a sum of $(k-1)(N-k+1)$ independent random variables, of which one of them is the contact rate λ_{k-1}^{next}, it follows that

$$E[S_{k-1}\cdot\lambda_{k-1}^{next}] = E[(\lambda_{k-1}^{next})^2] + [(k-1)(N-k+1)-1]\cdot\mu_{k-1}\cdot E[\lambda_{k-1}^{next}] \tag{2.71}$$

Substituting Eq. (2.71) in Eq. (2.70), and using the expression derived in Eq. (2.49), we get

$$\begin{aligned}
&Cov\left[S_{k-1},\lambda_{k-1}^{next}\right] = \\
&= E[(\lambda_{k-1}^{next})^2] + [(k-1)(N-k+1)-1]\cdot\mu_{k-1}\cdot E[\lambda_{k-1}^{next}] - (k-1)(N-k+1)\cdot\mu_{k-1}\cdot E[\lambda_{k-1}^{next}] \\
&= E[(\lambda_{k-1}^{next})^2] - \mu_{k-1}\cdot E[\lambda_{k-1}^{next}] \\
&= \left(Var\left[\lambda_{k-1}^{next}\right] + (E[\lambda_{k-1}^{next}])^2\right) - \mu_{k-1}\cdot E[\lambda_{k-1}^{next}] \\
&= Var\left[\lambda_{k-1}^{next}\right] + E[\lambda_{k-1}^{next}]\cdot\left(E[\lambda_{k-1}^{next}] - \mu_{k-1}\right)
\end{aligned} \tag{2.72}$$

Remark: In the previous derivations we used the the expression that relates the second moment of a random variable with its variance and mean value, i.e.

$$Var\left[X\right] = E[x^2] - (E[x])^2 \quad\Leftrightarrow\quad E[x^2] = Var\left[X\right] + (E[x])^2$$

Substituting in Eq. (2.72) the expression of Eq. (2.52), gives

$$Cov\left[S_{k-1},\lambda_{k-1}^{next}\right] = Var\left[\lambda_{k-1}^{next}\right] + \left(\mu_{k-1} + \epsilon_{k-1}^{next}\right)\cdot\epsilon_{k-1}^{next} \tag{2.73}$$

(v) We, similarly, express the second covariance appearing in the sum of Eq. (2.65) as

$$Cov\left[S_{k-1},S_{k-1}^{next}\right] = E[S_{k-1}\cdot S_{k-1}^{next}] - E[S_{k-1}]\cdot E[S_{k-1}^{next}] \tag{2.74}$$

The sum S_{k-1}^{next} consists of $k-2$ terms, which are also included in the sum S_{k-1}. For each term λ^* in S_{k-1}^{next}, there are $(k-1)(N-k+1)-1$ terms in S_{k-1} that are independent of λ^*. Hence, we can, successively, write for Eq. (2.74)

$$\begin{aligned} Cov\left[S_{k-1}, S_{k-1}^{next}\right] &= (k-2)\cdot E[S_{k-1}\cdot\lambda^*] - E[S_{k-1}]\cdot(k-2)\cdot E[\lambda^*] \\ &= (k-2)\left((\sigma_{k-1}^2+\mu_{k-1}^2) + [(k-1)(N-k+1)-1]\mu_{k-1}^2\right) \\ &\quad - (k-1)(N-k+1)\mu_{k-1}\cdot(k-2)\cdot\mu_{k-1} \\ &= (k-2)\cdot\sigma_{k-1}^2 \end{aligned} \tag{2.75}$$

Substituting in Eq. (2.65) the expressions we derived in (i)-(v) (Eq. (2.66), Eq. (2.68), Eq. (2.69), Eq. (2.73) and Eq. (2.75)), we get

$$\begin{aligned} Var\left[S_k\right] &= (k-1)(N-k+1)\cdot\sigma_{k-1}^2 + Var\left[\lambda_{k-1}^{next}\right] + (k-2)\cdot\sigma_{k-1}^2 \\ &\quad + (N-k)\cdot\sigma_\lambda^2 - 2\cdot\left(Var\left[\lambda_{k-1}^{next}\right] + \left(\mu_{k-1}+\epsilon_{k-1}^{next}\right)\cdot\epsilon_{k-1}^{next}\right) - 2\cdot(k-2)\cdot\sigma_{k-1}^2 \\ &= (k-1)(N-k+1)\cdot\sigma_{k-1}^2 - (k-2)\cdot\sigma_{k-1}^2 + (N-k)\cdot\sigma_\lambda^2 \\ &\quad - \left(Var\left[\lambda_{k-1}^{next}\right] + 2\cdot\left(\mu_{k-1}+\epsilon_{k-1}^{next}\right)\cdot\epsilon_{k-1}^{next}\right) \\ &= (k-1)(N-k)\sigma_{k-1}^2 + (N-k)\sigma_\lambda^2 - \left(Var\left[\lambda_{k-1}^{next}\right] + 2\cdot\left(\mu_{k-1}+\epsilon_{k-1}^{next}\right)\cdot\epsilon_{k-1}^{next} - \sigma_{k-1}^2\right) \\ &= k(N-k)\cdot\left[\frac{(k-1)\cdot\sigma_{k-1}^2+\sigma_\lambda^2}{k} - \frac{Var\left[\lambda_{k-1}^{next}\right] + 2\cdot\left(\mu_{k-1}+\epsilon_{k-1}^{next}\right)\cdot\epsilon_{k-1}^{next} - \sigma_{k-1}^2}{k(N-k)}\right] \end{aligned} \tag{2.76}$$

or

$$Var\left[S_k\right] = k(N-k)\cdot\left(\frac{(k-1)\cdot\sigma_{k-1}^2+\sigma_\lambda^2}{k} - \delta_k^{'}\right) \tag{2.77}$$

where

$$|\delta_k^{'}| = O\left(\frac{\lambda_{max}^2}{k(N-k)}\right) \tag{2.78}$$

Following a recursive procedure (for $k=1,2,...$) as previously, it can be shown that $\forall k$ it holds

$$Var\left[S_k\right] = k(N-k)\cdot\sigma_\lambda^2\cdot(1-\delta_k), \quad |\delta_k| = O\left(\frac{\lambda_{max}^2}{N-1}\right) \tag{2.79}$$

which proves the second part of Lemma 1. □

2.6.2 Proof of Lemma 2

Proof. Using Lemma 1 we can write for the random variable $X_k = \frac{S_k}{k(N-k)}$:

$$E[X_k] = \frac{E[S_k]}{k(N-k)} = \frac{k(N-k)\cdot\mu_\lambda(1-\epsilon_k)}{k(N-k)} = \mu_\lambda\cdot(1-\epsilon_k) \tag{2.80}$$

and

$$Var\left[X_k\right] = \frac{Var\left[S_k\right]}{\left(k(N-k)\right)^2} = \frac{k(N-k)\cdot\sigma_\lambda^2\cdot(1-\delta_k)}{\left(k(N-k)\right)^2} = \frac{\sigma_\lambda^2\cdot(1-\delta_k)}{k(N-k)} \tag{2.81}$$

The variance of X_k, Eq. (2.81), can be written as

$$Var\left[X_k\right] = E\left[\left(X_k - E[X_k]\right)^2\right] = E\left[\left(X_k - \mu_\lambda \cdot (1-\epsilon_k)\right)^2\right] = E\left[\left(X_k - \mu_\lambda\right)^2\right] - \left(\mu_\lambda \cdot \epsilon_k\right)^2 \tag{2.82}$$

where we used the expression of Eq. (2.80).

Combining Eq. (2.81) and Eq. (2.82), we get

$$E\left[\left(X_k - \mu_\lambda\right)^2\right] = \frac{\sigma_\lambda^2 \cdot (1-\delta_k)}{k(N-k)} + \left(\mu_\lambda \cdot \epsilon_k\right)^2 \tag{2.83}$$

and taking the limit, for $N \to \infty$, in both sides of Eq. (2.83), gives

$$\lim_{N\to\infty} E\left[\left(X_k - \mu_\lambda\right)^2\right] = \lim_{N\to\infty}\left(\frac{\sigma_\lambda^2 \cdot (1-\delta_k)}{k(N-k)} + \left(\mu_\lambda \cdot \epsilon_k\right)^2\right) = 0$$

since $\epsilon_k = O\left(\frac{\lambda_{max}}{N-1}\right)$.

Therefore, (by definition [70, Def. 5.3, p. 136]) it follows that

$$X_k \xrightarrow{m.s.} \mu_\lambda \tag{2.84}$$

where $\xrightarrow{m.s.}$ denotes convergence in mean square. □

2.6.3 Delivery Delay of Opportunistic Routing Protocols

2.6.3.1 Epidemic Routing

Substituting the expression for the expected step delay from Result 1 in Eq.(2.22) we get

$$E[T_D^{(epid)}] = \frac{1}{N-1}\sum_{k=1}^{N-1}(N-k)E[T_{k,k+1}] = \frac{\sum_{k=1}^{N-1}(N-k)\cdot\frac{1}{k(N-k)\mu_\lambda}\cdot\left(1+\frac{CV_\lambda^2}{[k(N-k)]}\right)}{N-1}$$

$$= \frac{1}{(N-1)\mu_\lambda}\sum_{k=1}^{N-1}\left(\frac{1}{k\mu_\lambda} + \frac{CV_\lambda^2}{k^2(N-k)}\right) = \frac{1}{(N-1)\mu_\lambda}\left[\sum_{k=1}^{N-1}\frac{1}{k} + CV_\lambda^2\sum_{k=1}^{N-1}\frac{1}{k^2(N-k)}\right] \tag{2.85}$$

Using *partial fraction decomposition*, the sum $\sum_{k=1}^{N-1}\frac{1}{k^2(N-k)}$ in Eq.(2.85) becomes

$$\sum_{k=1}^{N-1}\frac{1}{k^2(N-k)} = \frac{1}{N^2}\left(\sum_{k=1}^{N-1}\frac{1}{k} + \sum_{k=1}^{N-1}\frac{1}{N-k} + N\sum_{k=1}^{N-1}\frac{1}{k^2}\right). \tag{2.86}$$

We can approximate the *harmonic sum* $\sum_{k=1}^{N-1}\frac{1}{k}$ as [40]

$$\sum_{k=1}^{N-1}\frac{1}{k} \approx ln(N-1). \tag{2.87}$$

and, similarly, $\sum_{k=1}^{N-1}\frac{1}{k^2}$ can be approximated as [40]

$$\sum_{k=1}^{N-1}\frac{1}{k^2} \approx 1.65, \tag{2.88}$$

Using the approximations of Eq. (2.87) and Eq. (2.88) in Eq.(2.86) we get

$$\sum_{k=1}^{N-1} \frac{1}{k^2(N-k)} = \frac{1.65N + 2 \cdot ln(N-1)}{N^2}. \tag{2.89}$$

Substituting Eq.(2.87) and Eq.(2.89) in Eq.(2.85) and approximating $N-1 \approx N$ we get the expression in Table 2.3.

2.6.3.2 2-hop Routing

Under 2-hop routing, in *step* k there are k nodes that carry the message (the source and $k-1$ relays). As relays can forward the message only to the destination node and the source to everyone it meets, there are $N-1$ possible contact events in which a message exchange can take place, i.e. (a) $N-k-1$ possible meetings between the source and a non-infected node, other than the destination, and (b) k possible meetings between the infected nodes (including the source) and the destination. As a result, we can make the following two observations:

1. In contrast to epidemic spreading, where at each step k, the number of possible contact events is $k(N-k)$, here there are only $N-1$ possible contact events. Thus in the results we derived for the expected step delay $E[T_{k,k+1}]$, we should substitute the expression $k(N-k)$ with the expression $N-1$. For example, Result 1 in the case of *2-hop* routing becomes

$$E[T_{k,k+1}] \equiv E[T_{step}^{2hop}] \approx \frac{1}{(N-1)\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N-1}\right) \tag{2.90}$$

   As we can observe in Eq. (2.90), the expected step delay is independent of the step k, i.e. it is equal for every step k.

2. Due to randomness, the probability that the destination node will be involved in the exact next contact event with message exchange (conditioning that the message has not been delivered in any of the first $k-1$ steps) is

$$P\{\text{delivery at step } k | \text{no delivery before step } k\} \equiv P\{\text{at } k | \text{not before}\} = \frac{k}{(N-k-1)+k} = \frac{k}{N-1} \tag{2.91}$$

Then, from Eq. (2.91), it can be shown recursively that

$$P\{\text{delivery at step } k\} \equiv P\{\text{at } k\} = \frac{k}{N-1} \cdot \prod_{m=1}^{k-1} \left(1 - \frac{m}{N-1}\right) = \frac{k}{(N-1)^{k+1}} \cdot \frac{(N-1)!}{(N-k-1)!} \tag{2.92}$$

Now, if we denote as $E[T_D^{2hop}|\text{at } k]$ the expected delivery delay, given that the delivery takes place at step k, it holds that (using Eq. (2.90))

$$E[T_D^{2hop}|\text{at } k] = \sum_{m=1}^{k} E[T_{k,k+1}] = k \cdot E[T_{step}^{2hop}] \tag{2.93}$$

Therefore, using the property of conditional expectation, we can calculate the expected delivery delay for *2-hop* routing, as following:

$$\begin{aligned} E[T_D^{2hop}] &= \sum_{k=1}^{N-1} E[T_D^{2hop}|\text{at } k] \cdot P\{\text{at } k\} = \sum_{k=1}^{N-1} k \cdot E[T_{step}^{2hop}] \cdot P\{\text{at } k\} \\ &= \sum_{k=1}^{N-1} k \cdot E[T_{step}^{2hop}] \cdot \frac{k}{(N-1)^{k+1}} \cdot \frac{(N-1)!}{(N-k-1)!} \\ &\text{`} = E[T_{step}^{2hop}] \cdot \sum_{k=1}^{N-1} \frac{k^2}{(N-1)^{k+1}} \cdot \frac{(N-1)!}{(N-k-1)!} \\ &= \frac{1}{(N-1)\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N-1}\right) \sum_{k=1}^{N-1} \frac{k^2}{(N-1)^{k+1}} \cdot \frac{(N-1)!}{(N-k-1)!} \end{aligned} \tag{2.94}$$

where we used the expressions of Eq. (2.93) and Eq. (2.92).

We can further simplify Eq. (2.94) by using the approximation presented in [43]:

$$\begin{aligned} E[T_D^{2hop}] &= \frac{1}{(N-1)\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N-1}\right) \cdot \sum_{k=1}^{N-1} \frac{k^2}{(N-1)^{k+1}} \cdot \frac{(N-1)!}{(N-k-1)!} \\ &= \frac{1}{\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N-1}\right) \cdot \sum_{k=1}^{N-1} \frac{k^2}{(N-1)^{k+2}} \cdot \frac{(N-1)!}{(N-k-1)!} \approx \frac{1}{\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N-1}\right) \cdot \left[\frac{\sqrt{\frac{\pi}{2}}}{\sqrt{N}} + O\left(\frac{1}{N}\right)\right] \end{aligned} \tag{2.95}$$

2.6.3.3 Spray and Wait Routing

Among all the Spray and Wait (SnW) protocol versions, the one with the highest expected delivery delay is the *source-SnW* (i.e. where the source gives to each relay only 1 message copy) [129], which implies that $E[T_D^{(SnW)}] \leq E[T_D^{(source-SnW)}]$. Its mechanism is similar to the *2-hop* routing; the only difference is that now the number of message copies is limited, $L < (N-1)$.

Thus, the analysis follows similar steps. At first, for the expected step delay it holds

$$E[T_{k,k+1}] \approx \begin{cases} \frac{1}{(N-1)\cdot\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N-1}\right) & , \ k \leq L-1 \\ \frac{1}{L\cdot\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{L}\right) & , \ k = L \end{cases}$$

and for the probability of message delivery at step k, it holds that

$$P\{\text{at } k\} = \begin{cases} \frac{k}{(N-1)^{k+1}} \cdot \frac{(N-1)!}{(N-k-1)!} & , \ k \leq L-1 \\ \frac{1}{(N-1)^L} \cdot \frac{(N-1)!}{(N-L-1)!} & , \ k = L \end{cases} \tag{2.96}$$

Then, using the property of conditional expectation, we can calculate the expected delivery delay for *source-SnW* routing, as in Section 2.6.3.2.

2.6.4 Sketch of Proof of Corollary 1

As previously defined, $\mathbf{C}_k^m$ is the set of nodes with the message (the "infected" nodes) at step k. For each node $i \in \mathbf{C}_k^m$, we now define the set $\mathbf{D}_{\mathbf{C}_k^m}(i)$ as

$$\mathbf{D}_{\mathbf{C}_k^m}(i) = \{j : j \notin \mathbf{C}_k^m \; and \; \lambda_{ij} > 0\} \tag{2.97}$$

$\mathbf{D}_{\mathbf{C}_k^m}(i)$ is the set of the nodes j that have not received yet the message and can contact node i. In a full-mesh network (Def. 3), the cardinality of the set $\mathbf{D}_{\mathbf{C}_k^m}(i)$ is $\|\mathbf{D}_{\mathbf{C}_k^m}(i)\| = (N-k)$, whereas in a sparse network $0 \leq \|\mathbf{D}_{\mathbf{C}_k^m}(i)\| \leq N-k$. In particular, for the case we consider here (Poisson graphs), the sizes $\|\mathbf{D}_{\mathbf{C}_k^m}(i)\|$ are (approximately; and exactly in the limit of large N) binomially distributed[20] as

$$P\{\|\mathbf{D}_{\mathbf{C}_k^m}(i)\| = d\} = \binom{N-k}{d} \cdot (p_s)^d \cdot (1-p_s)^{(N-k)-d} \tag{2.98}$$

with

$$E\left[\|\mathbf{D}_{\mathbf{C}_k^m}(i)\|\right] = (N-k) \cdot p_s \tag{2.99}$$

$$Var\left[\|\mathbf{D}_{\mathbf{C}_k^m}(i)\|\right] = (N-k) \cdot p_s \cdot (1-p_s) \tag{2.100}$$

where the probability space is defined over all possible sets m at step k, and all nodes i.

Now, similarly to Eq. (2.4) and Eq. (2.5), we define

$$S_{k(p)}^m = \sum_{i \in \mathbf{C}_k^m} \sum_{j \in \mathbf{D}_{\mathbf{C}_k^m}(i)} \lambda_{ij} \tag{2.101}$$

and the random variable

$$P\{S_{k(p)} = S_k^m\} = P\{\mathbf{C}_k^m\} \tag{2.102}$$

In a full-mesh network, S_k^m is a sum of $k(N-k)$ terms λ_{ij}, and the moments of S_k are given by Lemma 1. In the Poisson graph case we consider here, $S_{k(p)}^m$ is a sum of $\sum_{i \in \mathbf{C}_k^m} \|\mathbf{D}_{\mathbf{C}_k^m}(i)\|$ terms, where the quantity

$$D_k^m = \sum_{i \in \mathbf{C}_k^m} \|\mathbf{D}_{\mathbf{C}_k^m}(i)\| \tag{2.103}$$

is a random variable as well.

Therefore, (i) taking into account that $S_{k(p)}^m$, as a sum of a random number (D_k^m) of i.i.d. random variables (λ_{ij})[21], (ii) making similar arguments as in the proof of Lemma 1, and (iii) neglecting terms $O\left(\frac{1}{N}\right)$ (see e.g. ϵ_k and δ_k in Lemma 1), it can be shown that the expectation and variance of $S_{k(p)}$

$$E[S_{k(p)}] = E[D_k^m] \cdot \mu_\lambda \tag{2.104}$$

$$Var\left[S_{k(p)}\right] = E[D_k^m] \cdot \sigma_\lambda^2 + \mu_\lambda^2 \cdot Var\left[D_k^m\right] \tag{2.105}$$

[20] This is because, by the definition of a Poisson graph, for each node $i \in \mathbf{C}_k^m$, there are $N-k$ other nodes $j \notin \mathbf{C}_k^m$, each of which is a neighbor of i with probability p_s and independently of all other links.

[21] Expressions for the statistic moments of sums of random number of random variables are given in [121].

Then, since D_k^m is a sum of k independent random variables (Eq. (2.103)), whose expectations and variances are given by Eq. (2.99) and Eq. (2.100), respectively, it follows that [121]

$$E[D_k^m] = k \cdot E\left[|\mathbf{D}_{\mathbf{C}_k^m}(i)\|\right] = k(N-k) \cdot p_s \tag{2.106}$$

$$Var\left[D_k^m\right] = k \cdot Var\left[|\mathbf{D}_{\mathbf{C}_k^m}(i)\|\right] = k(N-k) \cdot p_s \cdot (1-p_s) \tag{2.107}$$

Substituting Eq. (2.106) and Eq. (2.107) in the expressions of Eq. (2.104) and Eq. (2.105), we get

$$E[S_{k(p)}] = k(N-k) \cdot p_s \cdot \mu_\lambda \tag{2.108}$$

$$\begin{aligned} Var\left[S_{k(p)}\right] &= k(N-k) \cdot p_s \cdot \sigma_\lambda^2 + \mu_\lambda^2 \cdot k(N-k) \cdot p_s(1-p_s) \\ &= k(N-k) \cdot p_s \left[\sigma_\lambda^2 + \mu_\lambda^2 \cdot (1-p_s)\right] \end{aligned} \tag{2.109}$$

Comparing Eq. (2.108) and Eq. (2.109) to the corresponding expressions of Lemma 1, we can observe the correspondence suggested in Corollary 1.

2.6.5 Assuming a Constant Coefficient of Variation CV_d

From Eq. (2.28), we can easily result to the recurrence relation for the second moment of the degree distribution:

$$\overline{d^2}(k+1) = \frac{N-k}{N-(k+1)}\overline{d^2}(k) - \frac{1}{N-(k+1)}\frac{\overline{d^3}(k)}{\mu_d(k)} \tag{2.110}$$

where $\overline{d^n}(k)$ is the n^{th} moment of the degree distribution.

As we have computed the expectation and the second moment of the degree distribution in step $k+1$, Eq. (2.29) and Eq. (2.110) respectively, we can find the recurrence relation for the coefficient of variation, which is:

$$CV_d^2(k+1) = \frac{CV_d^2(k) \cdot \left(1 - \frac{\gamma_d(k) \cdot CV_d(k) + 2}{N-k-1}\right) + 1}{\left(1 - \frac{CV_d^2(k)}{N-k-1}\right)^2} - 1 \tag{2.111}$$

where we denote as $\gamma_d(k)$ the skewness of the degree distribution. In Eq. (2.111), if we do not know the value of $\gamma_d(k)$, we cannot solve the recurrence relation for $CV_d(k)$ and we cannot evaluate it. Thus, as we can see, the expression for the value of $CV_d^2(k)$ (which is equivalent to the second moment $E[d^2(k)]$) includes the value of the third moment of the degree distribution at state k. So, recursively, it follows that the exact solution of Eq. (2.29) requires the knowledge of all the higher moments of the degree distribution. However, this requirement both increases complexity and decreases applicability as it is not always efficient or possible to know or estimate all the higher moments of the degree distribution. Therefore, in order to find a closed form solution for $\mu_d(k)$, we assume $CV_d(k) = CV_d \ \ \forall k$.

This relation hold for the cases where $\frac{\gamma_d(k) \cdot CV_d(k) + 2}{N-k-1} \ll 1$ and $\frac{CV_d^2(k)}{N-k-1} \ll 1$, where it is easy to see from Eq. (2.111) that

$$CV_d^2(k+1) \simeq CV_d^2(k) \tag{2.112}$$

Summarizing, it is relatively accurate to assume that the coefficient of variation of the degree distribution remains the same for each state k, when

$$N - k \gg max\{1, \ \ CV_d^2, \ \ \gamma_d \cdot CV_d^2\} \tag{2.113}$$

2.6.6 Proof of Result 3

2.6.6.1 Derivation of the Recurrence Relation of Eq. (2.33)

At first we provide the derivation of the recurrence relation for the mean *out degree* in each step, i.e. $\overline{D}^{out}(k)$.

Proof. At step k, the average degree of the nodes that do not have the message is $\mu_d(k)$ and is given by Eq. (2.31). Thus, it holds that the total number of edges, connected to them, is $(N-k)\cdot\mu_d(k)$. Let the *out degree* to be $D^{out}(k)$ and the degree of the next node to receive the message to be $d^{new}(k)$ [22]. According to the reasoning of Section 2.3.2.1.2, the *out degree* of the next step will be

$$D^{out}(k+1) = D^{out}(k) + (d^{new}(k) - 2) - 2\cdot\mathcal{H}(M,m,n) \tag{2.114}$$

where $\mathcal{H}(M,m,n)$ is a random variable drawn from a *Hypergeometric* distribution [23] with parameters

$$\begin{aligned} M &= (N-k)\cdot\mu_d(k) - 1 \\ m &= d^{new}(k) - 1 \\ n &= D^{out}(k) - 1 \end{aligned}$$

Taking the expectation of both sides of Eq. (2.114) we get

$$\overline{D}^{out}(k+1) = \overline{D}^{out}(k) + (\mu_d^{new}(k) - 2) - 2\cdot E[\mathcal{H}(M,m,n)] \tag{2.115}$$

The value of $\mu_d^{new}(k)$ is given by Result 2. We cannot calculate directly the expectation of the Hypergeometric distribution, because its arguments are random variables too. Therefore, we need to compute first the conditional expectation, conditioning on $D^{out}(k)$ and $d^{new}(k)$:

$$\begin{aligned} E[\mathcal{H}(M,m,n)] &= \sum_{D^{out'}}\sum_{d^{new'}} E[\mathcal{H}(M,m,n)|\, D^{out'}, d^{new'}]\cdot P\left(D^{out'}, d^{new'}\right) \\ &= \sum_{D^{out'}}\sum_{d^{new'}} \frac{n\cdot m}{M}\cdot P\left(D^{out'}, d^{new'}\right) \\ &= \sum_{D^{out'}}\sum_{d^{new'}} \frac{(d^{new'}-1)\cdot(D^{out'}-1)}{(N-k)\cdot\mu_d(k)-1}\cdot P\left(D^{out'}, d^{new'}\right) \end{aligned} \tag{2.116}$$

and as $D^{out}(k)$ and $d^{new}(k)$ are independent random variables, then Eq. (2.116) becomes

$$E[\mathcal{H}(M,m,n)] = \frac{(\mu_d^{new}(k)-1)\cdot(D^{out}(k)-1)}{(N-k)\cdot\mu_d(k)-1} \tag{2.117}$$

and Eq. (2.115) turns into Eq. (2.33). □

[22] Note that $D^{out}(k)$ and $d^{new}(k)$ are not expectations, but they are random variables.

[23] The *Hypergeometric* distribution is a discrete probability distribution that describes the probability of l successes in n draws from a finite population of size M, containing m successes, *without* replacement.

2.6.6.2 Solution of Eq. (2.36)

Having derived the recurrence relation Eq. (2.33), we solve its equivalent expression, which is given by Eq. (2.36).

Proof. For $k = 1$, Eq. (2.36) gives:

$$\overline{D}^{out}(2) = \overline{D}^{out}(1) \cdot \left[1 - 2\frac{1 + CV_d^2}{N-1}\right] + (1 + CV_d^2) \cdot \mu_d(1),$$

for $k = 2$, it gives:

$$\begin{aligned}
&\overline{D}^{out}(3) = \overline{D}^{out}(2) \cdot \left[1 - 2\frac{1 + CV_d^2}{N-2}\right] + (1 + CV_d^2) \cdot \mu_d(2) \\
&= \overline{D}^{out}(1) \cdot \left[1 - 2\frac{1 + CV_d^2}{N-1}\right] \cdot \left[1 - 2\frac{1 + CV_d^2}{N-2}\right] + (1+CV_d^2) \cdot \mu_d(1) \cdot \left[1 - 2\frac{1 + CV_d^2}{N-2}\right] + (1+CV_d^2) \cdot \mu_d(2)
\end{aligned}$$

and recursively, it can be expressed as

$$\overline{D}^{out}(k) = \overline{D}^{out}(1) \cdot \prod_{m=1}^{k-1} \left[1 - 2\frac{1 + CV_d^2}{N-m}\right] + \sum_{m=1}^{k-1} (1 + CV_d^2) \cdot \mu_d(k) \prod_{\ell=m+1}^{k-1} \left[1 - 2\frac{1 + CV_d^2}{N-\ell}\right] \tag{2.118}$$

To find a closed-form expression for Eq. (2.118) we need first to calculate the sums and products separately. So, at first:

$$\begin{aligned}
\prod_{m=1}^{k-1} \left[1 - 2\frac{1 + CV_d^2}{N-m}\right] &\approx \prod_{m=1}^{k-1} e^{-2\frac{1+CV_d^2}{N-m}} \\
&= exp\left\{-2\left(1 + CV_d^2\right) \cdot \sum_{m=1}^{k-1} \frac{1}{N-m}\right\} \\
&= exp\left\{-2\left(1 + CV_d^2\right) \cdot \sum_{m=N-k+1}^{N-1} \frac{1}{m}\right\} \\
&\approx exp\left\{-2\left(1 + CV_d^2\right) \cdot [\ln(N-1) - \ln(N-k)]\right\} \\
&= exp\left\{-2\left(1 + CV_d^2\right) \cdot \ln\left(\frac{N-1}{N-k}\right)\right\} \\
&= \left(\frac{N-k}{N-1}\right)^{2\left(1+CV_d^2\right)} \\
&= \left(\frac{N-k}{N-1}\right) \cdot \left(\frac{N-k}{N-1}\right)^{1+2CV_d^2}
\end{aligned} \tag{2.119}$$

where for the first approximation we used the Taylor series expansion (similarly to the proof of Result 2), which is accurate for $N - k > 4(1 + CV_d^2)$, and for the second approximation we used the harmonic series approximation, whose accuracy increases for larger values of $N - k$.

Similarly to Eq. (2.119), we can find that

$$\prod_{\ell=m+1}^{k-1}\left[1-2\frac{1+CV_d^2}{N-\ell}\right]\approx\left(\frac{N-k}{N-m-1}\right)^{2\left(1+CV_d^2\right)} \tag{2.120}$$

and now, using Eq. (2.120), we can write for the summation in Eq. (2.118)

$$\begin{aligned}\sum_{m=1}^{k-1}(1+CV_d^2)\cdot\mu_d(k)\prod_{\ell=m+1}^{k-1}\left[1-2\frac{1+CV_d^2}{N-\ell}\right]&\\ =\sum_{m=1}^{k-1}(1+CV_d^2)\cdot\mu_d(k)\cdot\left(\frac{N-k}{N-m-1}\right)^{2\left(1+CV_d^2\right)}&\\ =(1+CV_d^2)\sum_{m=1}^{k-1}\mu_d\left(\frac{N-m-1}{N-k}\right)^{CV_d^2}\cdot\left(\frac{N-k}{N-m-1}\right)^{2\left(1+CV_d^2\right)}&\\ =(1+CV_d^2)\cdot\mu_d\cdot\frac{(N-k)^{2\left(1+CV_d^2\right)}}{(N-1)^{CV_d^2}}\cdot\sum_{m=1}^{k-1}\left(\frac{1}{N-m-1}\right)^{2+CV_d^2}&\\ =(1+CV_d^2)\cdot\mu_d\cdot\frac{(N-k)^{2\left(1+CV_d^2\right)}}{(N-1)^{CV_d^2}}\cdot\sum_{m=N-k}^{N-2}\frac{1}{m^{2+CV_d^2}}&\end{aligned} \tag{2.121}$$

We approximate the sum that appears in the right side of the last line in Eq. (2.121) with the integral

$$\begin{aligned}\sum_{m=N-k}^{N-2}\frac{1}{m^{2+CV_d^2}}&\approx\int_{N-k}^{N-1}\frac{1}{m^{2+CV_d^2}}dm\\ &=\frac{(N-1)^{\left(1-(2+CV_d^2)\right)}-(N-k)^{\left(1-(2+CV_d^2)\right)}}{1-(2+CV_d^2)}\\ &=\frac{1}{1+CV_d^2}\left[\frac{1}{(N-k)^{1+CV_d^2}}-\frac{1}{(N-1)^{1+CV_d^2}}\right]\end{aligned} \tag{2.122}$$

and finally, combining Eq. (2.121) and Eq. (2.122), we get

$$\begin{aligned}\sum_{m=1}^{k-1}(1+CV_d^2)\cdot\mu_d(k)\prod_{\ell=m+1}^{k-1}\left[1-2\frac{1+CV_d^2}{N-\ell}\right]&\\ =\mu_d\cdot\frac{(N-k)^{2\left(1+CV_d^2\right)}}{(N-1)^{CV_d^2}}\cdot\left[\frac{1}{(N-k)^{1+CV_d^2}}-\frac{1}{(N-1)^{1+CV_d^2}}\right]&\\ =\mu_d\cdot(N-k)\cdot\left[\left(\frac{N-k}{N-1}\right)^{CV_d^2}-\left(\frac{N-k}{N-1}\right)^{1+2CV_d^2}\right]&\end{aligned} \tag{2.123}$$

Substituting in Eq.(2.118) the expressions from Eq.(2.119) and Eq.(2.123) and having in

mind that $\overline{D}^{out}(1) = \mu_d$, we can write

$$\begin{aligned}\overline{D}^{out}(k) &= \mu_d \cdot \left(\frac{N-k}{N-1}\right) \cdot \left(\frac{N-k}{N-1}\right)^{1+2CV_d^2} \\ &+ \mu_d \cdot (N-k) \cdot \left[\left(\frac{N-k}{N-1}\right)^{CV_d^2} - \left(\frac{N-k}{N-1}\right)^{1+2CV_d^2}\right] \\ &= \mu_d \cdot (N-k)\left[\left(\frac{N-k}{N-1}\right)^{CV_d^2} - \left(1-\frac{1}{N-1}\right)\left(\frac{N-k}{N-1}\right)^{1+2CV_d^2}\right] \end{aligned} \tag{2.124}$$

which gives Result 3. □

2.6.7 Proof of Result 4

Applying the condition $\mu_d(k) \geq d_{min}$ in Eq. (2.31), we can find that it is satisfied for the steps k that

$$k \leq \left[1 - \left(\frac{d_{min}}{\mu_d}\right)^{\frac{1}{CV_d^2}}\right] \cdot (N-1) = k_{stop} \tag{2.125}$$

The previous equation means that after the k_{stop}^{th} state [24], Eq. (2.31), gives values $\mu_d(k) \leq d_{min}$. To overcome this problem, we will use Eq. (2.32) for calculating $\overline{D}^{out}(k)$ for $k \leq k_{stop}$ till step k_{stop} and then, as all the remaining nodes must have degree d_{min}, use the recurrence relation:

$$\overline{D}^{out}(k+1) = \overline{D}^{out}(k) + d_{min} - \frac{2 \cdot \overline{D}^{out}(k)}{N-k} \tag{2.126}$$

Solving, similarly as in Appendix 2.6.6, the Eq. (2.126), for initial condition $\overline{D}^{out}(k_{stop}) = D_{stop}$ where the value of D_{stop} is taken from Eq. (2.32), we end up to the recurrence relation

$$\overline{D}^{out}(k) = D_{stop} \cdot \prod_{m=k_{stop}}^{k-1}\left(1 - \frac{2}{N-m}\right) + d_{min} \cdot \sum_{m=k_{stop}}^{k-1}\prod_{\ell=m+1}^{k-1}\left(1-\frac{2}{N-m}\right) \tag{2.127}$$

for $k > k_{stop}$. Using the Taylor series expansion and Harmonic series approximations we can show that

$$\prod_{m=k_{stop}}^{k-1}\left(1-\frac{2}{N-m}\right) = \left(\frac{N-k}{N-k_{stop}}\right)^2 \tag{2.128}$$

$$\prod_{\ell=m+1}^{k-1}\left(1-\frac{2}{N-m}\right) = \left(\frac{N-k}{N-m-1}\right)^2 \tag{2.129}$$

[24] In case $k_{stop} > N-1$ the following analysis is not needed and we can use the Result 3

and then, by Eq. (2.129):

$$\begin{aligned}\sum_{m=k_{stop}}^{k-1}\prod_{\ell=m+1}^{k-1}\left(1-\frac{2}{N-m}\right) &= \sum_{m=k_{stop}}^{k-1}\left(\frac{N-k}{N-m-1}\right)^2 \\ &= (N-k)^2\cdot\sum_{m=N-k}^{N-k_{stop}-1}\frac{1}{m^2} \\ &\approx (N-k)^2\cdot\int_{m=N-k}^{N-k_{stop}}\frac{1}{m^2}dm \\ &= (N-k)^2\cdot\left[\frac{1}{N-k}-\frac{1}{N-k_{stop}}\right] \qquad (2.130)\end{aligned}$$

Now, Result 4 follows easily by substituting the expressions of Eq. (2.128) and Eq. (2.130) in Eq. (2.127).

Remark: As we saw, in our analysis, we first consider Result 2 and for the last steps we assume $\mu_d^{new}(k) = d_{min}$ in order to derive Result 4. In addition to the intuitive reasons, which we described, this assumption can also justified by a similar work. In [7], the authors investigate, through analysis and simulations, the average degree of the newly infected nodes, $\mu_d^{new}(k)$. They conclude that in early steps $\mu_d^{new}(k)$ is given by $\frac{\overline{d^2}}{\overline{d}} = \mu_d\cdot(1+CV_d^2)$, which is in agreement with our result, and then it gradually decreases and in the last steps it becomes equal to the minimum degree of the network, d_{min}.

Chapter 3

Understanding the Effects of Social Selfishness

3.1 Introduction

An abundance of routing techniques have been proposed for MSNs, comprising social-oblivious protocols (e.g. epidemic [137], two-hop [43], spray and wait [129]) where replication is used as a diversity mechanism to improve performance, and social-aware protocols (see [146] for a survey) where "good" relays are selected based on social (or other) characteristics. It is common for these protocols to be extensively evaluated under various mobility environments, using synthetic simulators, mobility models, or real mobility traces. There exist also a number of studies comparing the performance tradeoffs (e.g. delivery delay or probability vs. number of transmissions per message, etc.) of different protocols in different mobility environments.

However, the vast majority of works proposing, modeling, or optimizing protocols for MSNs assume cooperation of nodes in relaying messages: when the protocol dictates that a relay node should receive or transmit a message (neither destined to nor originating from it), it does. In practice, a relay node might: (i) never be willing to carry traffic for 3^{rd} parties, (ii) be willing to only perform some number of transmissions/receptions for relay traffic, or (iii) be more willing to receive or transmit traffic from nodes it has some (social) "ties" with. The reasons for this reluctance range from privacy concerns (e.g. not trusting an exchange with unknown nodes) to resource consumption (e.g. battery depletion). Such behaviors are natural, and could significantly degrade the predicted performance of the above protocols.

To this end, some recent works have used both simulations and analysis to study the effect of having some "selfish" nodes among "altruistic" nodes [61], or the effect of nodes reducing the transmission probability for all relay traffic (e.g. accepting or forwarding a packet with a probability $p < 1$) [88, 106]. Nevertheless, these works assume a mostly *uniform* behavior of relays when it comes to treating contacts with different nodes.

Contrary to the above approach, everyday experience suggests that people take into account the strength of their relation with a peer, when deciding whether to cooperate or not (*social selfishness*). As a result, a node A may be more willing to spend some energy (or take the risk) to forward a message of possible interest to an encountered node B, if A and B have strong *ties*, than if B is unknown to A. Furthermore, a long line of research has revealed that: (i) the strength of the "social" tie between two nodes (where "social" here may also be context-dependent) can

often be reasonably predicted by the contact rate between them [55, 108], and (ii) the contact patterns and rates between mobile nodes exhibit significant amounts of heterogeneity [25, 36].

This opens up a very large space of possible cooperation policies, whose performance might be intimately related to the underlying mobility. E.g. a node might choose to forward (or accept) messages only to (from) nodes that it encounters frequently enough, or attempt to explore "weak ties" [64]. Alternatively, a node could instead "modulate" the forwarding probability as a function of the encounter rate with a given node. The following questions are then raised:

Q.1 Can we predict the performance of a routing mechanism, under a given cooperation policy, if we only know some basic statistics about the underlying heterogeneous mobility process?

Q.2 Can we improve performance by choosing the cooperation policy wisely, subject to a given constraint (e.g. power consumption rate for relay traffic)?

The former question is relevant, for example, when the policy is given (related to external, e.g. security factors). One then might like to know what kind of performance he should expect from the network, so as to choose the right protocol or protocol parameters, without knowing the global network topology, or to decide whether opportunistic networking is useful in this context or it is better to simply use the infrastructure. The latter question is relevant when we can assume that the average node is willing to contribute some fixed amount of resources (e.g. amount of power spent for relay traffic) towards participating in an opportunistic network, but we are interested in how to best use these resources to optimize network performance.

Our main contributions in this chapter are

- We propose a generic model for *social selfishness* (or cooperation) related to mobility, which can capture a wide range of selfish behaviors and describe cooperation policies proposed in past literature (Section 3.2).

- Towards answering the first question, we use our model to provide closed-form expressions for the expected message delivery delay under a large class of mobility scenarios with heterogeneous contact rates; these expressions provide insights about the effect of the cooperation policy used and of the macroscopic mobility properties (mean value and variance of contact rates) (Section 3.3).

- Towards answering the second question, we examine the achievable performance-power consumption tradeoff *regions* under different cooperation policies. Specifically, we show that (i) when considering an interesting class of *Power-vs-Delay* tradeoffs, complex "social-based" policies cannot achieve better performance than the simple uniform policy, while (ii) when we consider *Power-vs-Delivery-Probability* tradeoffs, social cooperation policies can indeed be optimized (Section 3.4).

- Finally, we show that the intuition of our framework can be useful also in some real-world scenarios with significantly more complexity than the class of heterogeneous mobility models that we consider for our analysis.

3.2 Social Selfishness Models

We consider a network $\mathcal{N}$, with N nodes. We assume that nodes contact each other according to the Heterogeneous Contact Network model of Def. 3[1]. Users exchange messages using the *store-carry-forward* paradigm (like the epidemic-based protocols described in Chapter 2), and communication can be end-to-end (e.g. unicast) or content-centric.

The *store-carry-forward* mechanism requires from the relay nodes to (i) receive messages, (ii) store them, (iii) forward the messages they have to other relays and/or the destinations. As it is evident, this mechanism requires the cooperation of the relay nodes, and may put a heavy toll on their resources (bandwidth, storage space, battery life, etc.), dependent on the network traffic and protocol used. Furthermore, exchanging messages with unknown nodes may raise important security and privacy concerns. These considerations may render wireless nodes reasonably reluctant to relay traffic.

This unwillingness to cooperate might come in different flavors:
1. A node will not relay any traffic (*individual selfishness*).
2. A node will choose to relay each packet with some probability p. We will call this *uniform selfishness*.
3. A node will relay packets preferentially to other nodes it has a social relationship with (*social selfishness*).

The first case is an extreme case that could be handled with incentive or reputation mechanisms [22, 74, 94, 127, 145]. Such mechanisms are orthogonal (but possibly complementary) to our work. The second scenario has already been addressed in the past, with both theory and simulations, indicating that a low p can significantly hurt performance (e.g. [106]). The third case is closer, in our opinion, to human behavior. It is reasonable to assume that nodes are more willing to forward messages to or receive messages from nodes with whom they have a social tie. A social tie can be considered as a social relation in the real world (e.g. friendship), as a relation that originates from a routing mechanism (e.g. common interests in social-aware routing, SANE [92]), as a trust-relation that depends on how many times they have met in the past or they have collaborated (e.g. message exchanges or participation in a service composition) etc.

An important observation (for opportunistic networking) is that such social ties seem to be related with the mobility patterns. Studies from sociology [41] and social media [37] have shown that the stronger the social tie between two people is, the more they tend to meet or contact each other. Another study of Social Pervasive Networks [108], based on results from the anthropology field [144], shown that a relation between social ties and contact frequency (e.g. interaction on the respective social network) is supported in real networks. More recently, studies have directly suggested that the actual physical contact (related to mobility) can often serve as a good predictor for the strength of a social tie [54].

Combining the relations, we discussed above, between (i) *selfishness* and *social ties*, and (ii) *social ties* and *mobility patterns*, it is reasonable to assume a social selfishness model, where nodes decide to utilize a given contact opportunity with a probability $p_{ij} = p(\lambda_{ij})$, related to the contact frequency between the two nodes involved $\{i, j\}$. Such a model has been taken into account in a number of studies of routing protocols or message dissemination performance [64, 82, 84]. Some proposed strategies are, for example, to give more emphasis to "strong ties" or "weak ties" (i.e. large or small λ_{ij}): e.g. a node might decide to exchange messages only with

[1]The analysis and results of this chapter apply to the corresponding sparse network models of Section 2.3 (considering though heterogeneous rates λ_{ij}) as well.

the nodes it contacts more frequently, or the probability for message exchange to be linearly increasing with the contact rate of a pair of nodes, etc.

To be able to capture most of the above selfishness behaviors (and more), in a simple and generic way, we choose to model this willingness to forward a message (essentially, the existence of related constraints affecting this willingness), in a *probabilistic* way.

Specifically, we propose two types of selfishness models, which correspond to typical behaviors that can appear in a MSN.

Definition 7. *[Selfishness: Type I] The probability for a message to be exchanged in a contact event between two nodes i and j, depends on their meeting rate λ_{ij} and is described by the relation:*

$$p_{ij} = p^{(I)}(\lambda_{ij}), \qquad p_{ij} \in [0,1] \tag{3.1}$$

Definition 8. *[Selfishness: Type II] A pair of nodes i and j either can exchange messages in every contact event with probability p_{ij} or can never exchange messages with probability $1 - p_{ij}$. The probability p_{ij} depends on the meeting rate between these nodes, i.e. λ_{ij}, and is described by the relation:*

$$p_{ij} = p^{(II)}(\lambda_{ij}), \qquad p_{ij} \in [0,1] \tag{3.2}$$

The probabilities for message exchange may depend, as described earlier, on various factors, e.g. willingness of the nodes, routing protocol mechanism, battery constraints, duration of the contact. The above two models allows to capture a number of such concerns. Furthermore, Type II selfishness is useful to capture situations where nodes decide a priori whether they will interact with a given node or not (e.g. due to security concerns), while Type I selfishness models situations where the contact probability might be modulated according, for example, to current battery level, content sensitivity, desire to control relay traffic, etc.

3.3 Message Delivery Delay

Having defined the types of node mobility and the types of node selfishness that we consider, we can now commence our analysis. Our goal is twofold:

1) To capture the combined effect of all nodes applying a given "selfishness" policy (or cooperation policy, to be less negative) on the performance of basic opportunistic routing protocols (e.g. epidemic routing, spray and wait, etc.).

2) To compare different cooperation behaviors and understand the impact of mobility properties on absolute and relative performance.

We state upfront that an *exact* analysis of random opportunistic routing protocols is already very challenging for Heterogeneous Contact Networks (as explained in Section 2.2), and it becomes significantly more complex when social-selfishness policies are considered. For this reason, we try instead to derive useful closed form approximations, that can be directly used for performance predictions as well as policy optimization.

3.3.1 Effect of Social Selfishness

In Result 1 we shown (using the *Delta method* [103]) that the expectation of the spreading delay from state k to state $k+1$ for a Heterogeneous Contact Network can be approximated

with a series expansion as

$$E[T_{k,k+1}] = \frac{1}{M \cdot \mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{M} + R\right) = \frac{1}{M \cdot \mu_\lambda} + \frac{CV_\lambda^2}{M^2} + R \tag{3.3}$$

where $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$, $M = k \cdot (N-k)$ when epidemic routing is considered, and $R = \mathcal{O}(\frac{1}{M^2})$ corresponds to the impact of higher order terms.

The quantity M denotes the number of *eligible* pairs $\{i, j\}$ of nodes whose contact will take the process to the next state. In epidemic routing, each pair of nodes $\{i, j\}$, where i has the message and j does not, is an eligible pair. Under other epidemic-based schemes, not all such pairs are eligible. For instance, it is easy to see that under Spray and Wait routing with k copies, the number of eligible pairs (in the wait phase) is $M = k$. Therefore, generalizing Result 1[2], we can use Eq. (3.3) as an approximation for the step delay of different epidemic-based protocols by selecting appropriately the value of M.

When we introduce (social) selfishness, not all contacts resulting from the mobility model are useful in the spreading process, as was the case above. For instance, a node pair $\{i, j\}$ that meets with rate λ_{ij}, may exchange messages, on average, only half of the times (due to a Type I policy). Then, the *effective* (i.e. useful) contact rate will be $\lambda_{ij}^{'} = 0.5 \cdot \lambda_{ij}$.

The following lemmas give the mean value and variance of the *effective* contact rates in networks with contact rate probability function f_λ (with μ_λ and σ_λ^2) and selfishness of Type I (Lemma 3) or Type II (Lemma 4).

Lemma 3. *The mean value, $\mu_\lambda^{(I)}$, and the variance, $\sigma_\lambda^{2(I)}$, of the effective contact rates in a network with contact rate probability function f_λ (μ_λ, σ_λ^2) and selfishness of Type I, are given by*

$$\mu_\lambda^{(I)} = E\left[\lambda \cdot p^{(I)}(\lambda)\right] \tag{3.4}$$

$$\sigma_\lambda^{2(I)} = E\left[\lambda^2 \cdot \left(p^{(I)}(\lambda)\right)^2\right] - \left(E\left[\lambda \cdot p^{(I)}(\lambda)\right]\right)^2 \tag{3.5}$$

where the expectations are taken over the p.d.f. f_λ.

Proof. As defined in Def. 3, the contact process for a pair $\{i, j\}$ is a Poisson process with rate λ_{ij}. Thus, if, according to Def. 7, in each of the contact events a message can be exchanged with probability p_{ij} (independently of what happened in the previous or following contact events), then the *effective contact events* are described by another Poisson process, which results after thinning the initial contact process. The rate of the new, thinned, Poisson process is then

$$\lambda_{ij}^{(I)} = \lambda_{ij} \cdot p_{ij} = \lambda_{ij} \cdot p^{(I)}(\lambda_{ij})$$

Hence, the mean value of the rate of the effective contact events, is given by

$$\mu_\lambda^{(I)} = E\left[\lambda^{(I)}\right] = \int_0^\infty E\left[\lambda^{(I)} | \lambda_{ij} = x\right] \cdot f_\lambda(x)dx = \int_0^\infty \left(x \cdot p^{(I)}(x)\right) \cdot f_\lambda(x)dx = E\left[\lambda \cdot p^{(I)}(\lambda)\right]$$

[2]The same generalization has been used for deriving the results of Section 2.2.5.

Similarly the second moment, is given by

$$E\left[\left(\lambda^{(I)}\right)^2\right] = \int_0^\infty E\left[\left(\lambda^{(I)}\right)^2 |\lambda_{ij} = x\right] \cdot f_\lambda(x)dx = \int_0^\infty \left(x \cdot p^{(I)}(x)\right)^2 \cdot f_\lambda(x)dx = E\left[\lambda^2 \cdot \left(p^{(I)}(\lambda)\right)^2\right]$$

and, finally, the variance can be computed as:

$$\sigma_\lambda^{2(I)} = E\left[\left(\lambda^{(I)}\right)^2\right] - \left(\mu_\lambda^{(I)}\right)^2 = E\left[\lambda^2 \cdot \left(p^{(I)}(\lambda)\right)^2\right] - \left(E\left[\lambda \cdot p^{(I)}(\lambda)\right]\right)^2$$

□

Lemma 4. *The mean value, $\mu_\lambda^{(II)}$, and the variance, $\sigma_\lambda^{2(II)}$, of the effective contact rates in a network with contact rate probability function f_λ (μ_λ, σ_λ^2) and selfishness of Type II, are given by*

$$\mu_\lambda^{(II)} = E[\lambda \cdot p^{(II)}(\lambda)] \tag{3.6}$$

$$\sigma_\lambda^{2(II)} = E[\lambda^2 \cdot p^{(II)}(\lambda)] - \left(E[\lambda \cdot p^{(II)}(\lambda)]\right)^2 \tag{3.7}$$

where the expectations are taken over the p.d.f. f_λ.

Proof. According to Def. 8, a pair of nodes $\{i, j\}$ that contacts with rate λ_{ij}, either can always exchange a message during a contact event, with probability $p_{ij} = p^{(II)}(\lambda_{ij})$, or never exchanges messages during its contact events, with probability $1 - p_{ij}$. The equivalent of this constraint mechanism, is a network where some pairs of nodes contact with their initial rate, i.e. $\lambda_{ij}^{(II)} = \lambda_{ij}$, and some never contact, i.e. $\lambda_{ij}^{(II)} = 0$.

Thus, we can compute the mean value of the *effective* contact events as following:

$$\begin{aligned}\mu_\lambda^{(II)} &= \int_0^\infty E\left[\lambda^{(II)} | \lambda_{ij} = x\right] \cdot f_\lambda(x)dx \\ &= \int_0^\infty \left(x \cdot p^{(II)}(x) + 0 \cdot (1 - p^{(II)}(x))\right) \cdot f_\lambda(x)dx = \int_0^\infty \left(x \cdot p^{(II)}(x)\right) \cdot f_\lambda(x)dx = E[\lambda \cdot p^{(II)}(x)]\end{aligned}$$

Similarly,

$$\begin{aligned}&E\left[\left(\lambda^{(II)}\right)^2\right] = \int_0^\infty E\left[\left(\lambda^{(II)}\right)^2 |\lambda_{ij} = x\right] \cdot f_\lambda(x)dx \\ &= \int_0^\infty \left(x^2 \cdot p^{(II)}(x) + 0^2 \cdot (1 - p^{(II)}(x))\right) \cdot f_\lambda(x)dx = \int_0^\infty \left(x^2 \cdot p^{(II)}(x)\right) \cdot f_\lambda(x)dx = E[\lambda^2 \cdot p^{(II)}(x)]\end{aligned}$$

and finally

$$\sigma_\lambda^{2(II)} = E\left[\left(\lambda^{(II)}\right)^2\right] - \left(\mu_\lambda^{(II)}\right)^2 = E\left[\lambda^2 \cdot p^{(II)}(\lambda)\right] - \left(E\left[\lambda \cdot p^{(II)}(\lambda)\right]\right)^2$$

□

Thus, when the network is characterised by social selfishness, we can use the above expressions in Eq. (3.3) to calculate the delay $E[T_{k,k+1}]$.

As discussed earlier, one can use one, two or more terms of Eq. (3.3) (and the respective moments, e.g. $\mu_\lambda^{(I)}$, $\sigma_\lambda^{2(I)}$, etc.) to increase accuracy. However, by including many terms, expressions get complex and it might be difficult to be used for optimization or to provide insights. Thus, without loss of generality and in order to simplify our discussion, in the remainder we will use the simplest first order approximation, i.e. $E[T_{k,k+1}] = \frac{1}{M \cdot \mu_\lambda^{(I)}}$ for Type I selfishness (similarly for Type II).

Having computed the delay $E[T_{k,k+1}]$, we can now use the linearity of expectation rule to calculate the expected message *delivery* delay under different random routing protocols.

Result 5. *The expected message delivery delay in an Heterogeneous Contact Network can be approximated by*

$$E[T_D] = \frac{c(N,L)}{\mu_\lambda^{eff.}}, \tag{3.8}$$

where $\mu_\lambda^{eff.}$ *is given by Eq.* (3.4) *or Eq.* (3.6) *for selfishness of Type I or Type II, respectively, and* $c(N,L)$ *is a constant dependent on the size of the network,* N*, the routing protocol* $\mathcal{P}$ *and the number of message copies,* L*. Values of* $c(N,L)$ *are given in Table 3.1 for three well-known routing protocols*[3]*.*

In other words, as a first order approximation, the message delivery delay under random routing protocols is inversely proportional to the mean value of the effective contact rates in the network. Furthermore, the effect of Type I and Type II policies, with the same function $p(\lambda)$, turns out to be equal. We will thus not differentiate between the two policies, in the remainder, and simply refer to the mean effective contact rate as μ_λ^{eff}.

Finally, it is interesting to note that the effect of the mobility heterogeneity, in this first order approximation, when nodes are *not* selfish, affects performance only through its mean and not its variance (we have confirmed this to be the case for large N and non-heavy-tailed f_λ). In contrast, as we will show in the following sections, this is not the case when we introduce social selfishness in the spreading process.

Table 3.1: The values of $c(N,L)$ for three routing protocols.

Epidemic	$c(N,L) \approx \frac{\ln(N)}{N}$
2-hop	$c(N,L) = \sum_{k=1}^{N-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+2} \cdot (N-k-1)!}$
SnW	$c(N,L) \leq \sum_{k=1}^{L-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+2} \cdot (N-k-1)!} + \left(\frac{L}{N-1} + \frac{1}{L}\right) \frac{(N-1)!}{(N-1)^L \cdot (N-L-1)!}$

[3]The expressions in Table 3.1 are derived similarly to the corresponding expressions of Section 2.2.5.

Table 3.2: Mean effective contact rate, $\mu_\lambda^{eff.} = E[\lambda \cdot p(\lambda)]$.

Policy A	Policy B	Policy C	Policy D
Gamma: $f_\lambda(x) = \frac{\beta^a}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta x}, \quad \alpha, \beta, x > 0$		$\left(\text{where} \quad \alpha = \frac{1}{CV_\lambda^2}, \quad \beta = \frac{1}{\mu_\lambda \cdot CV_\lambda^2}\right)$	
$\mu_\lambda \cdot p_0$	$\mu_\lambda \cdot \left(p_2 + (p_1 - p_2)\frac{\gamma\left(1 + \frac{1}{CV_\lambda^2}, \frac{\lambda_0}{\mu_\lambda CV_\lambda^2}\right)}{\Gamma\left(1 + \frac{1}{CV_\lambda^2}\right)}\right)$	$\mu_\lambda \cdot p_1 \cdot \frac{\gamma\left(1 + \frac{1}{CV_\lambda^2}, \frac{\lambda_0}{\mu_\lambda CV_\lambda^2}\right)}{\Gamma\left(1 + \frac{1}{CV_\lambda^2}\right)} + p_2 \cdot \lambda_0 \cdot p_0$	$\mu_\lambda \cdot p_0 \left(1 - \frac{1}{\left(1 + m \cdot \mu_\lambda CV_\lambda^2\right)^{1 + \frac{1}{CV_\lambda^2}}}\right)$
	where $\quad \gamma\left(\frac{1}{CV_\lambda^2}, \frac{\lambda_0}{\mu_\lambda CV_\lambda^2}\right) = \Gamma\left(\frac{1}{CV_\lambda^2}\right) \cdot (1 - p_0)$		
Exponential: $f_\lambda(x) = \frac{1}{\mu_\lambda} \cdot e^{-\frac{x}{\mu_\lambda}}, \quad \mu_\lambda, x > 0$			
$\mu_\lambda \cdot p_0$	$\mu_\lambda \cdot \left(p_1 + (p_2 - p_1) \cdot p_0 \cdot [1 - \ln(p_0)]\right)$	$\mu_\lambda \cdot \left(p_1 - p_0 \cdot \left(p_1 - (p_2 - p_1) \cdot \ln(p_0)\right)\right)$	$\mu_\lambda \cdot p_0 \cdot \left(1 - \frac{1}{(1 + m \cdot \mu_\lambda)^2}\right)$
Pareto: $\overline{F}_\lambda(x) = \left(\frac{\beta}{x + \beta}\right)^\alpha, \quad \alpha, \beta, x > 0$		$\left(\text{where} \quad \alpha = \frac{2 \cdot CV_\lambda^2}{CV_\lambda^2 - 1}, \quad \beta = \mu_\lambda \cdot \frac{CV_\lambda^2 + 1}{CV_\lambda^2 - 1}\right)$	
$\mu_\lambda \cdot p_0$	$\mu_\lambda \cdot \left(p_1 + (p_2 - p_1) \cdot p_0 \cdot \left(1 - \alpha + \frac{\alpha}{p_0^{\frac{1}{\alpha}}}\right)\right)$	$\mu_\lambda \cdot \left(p_1 \cdot \left(1 - p_0 \cdot \left(1 - \alpha + \frac{\alpha}{p_0^{\frac{1}{\alpha}}}\right)\right) + p_2 \cdot p_0 \cdot \beta \cdot \left(\frac{1}{p_0^{\frac{1}{\alpha}}} - 1\right)\right)$	N.A.

3.3.2 Case Studies

With the basic performance result now in hand, we can go ahead and consider specific mobility processes, f_λ, and selfishness policies, $p(\lambda)$. To this end, we have analysed four policies (Table 3.3), which can represent a wide (and diverse) set of common behaviors for social selfishness and/or have been proposed before [64]. We will describe these policies only as Type I selfishness, but the analysis holds for the respective Type II policies as well.

Policy A *Uniform*: Each pair of nodes exchanges messages with probability p_0 every time they contact. The selfishness is not related with the contact rates between nodes.

Policy B *Strong / Weak ties*: Each pair of nodes exchanges messages with probability p_1 if they contact with rate less than λ_0 and with probability p_2 otherwise. The values of p_1 and p_2 determine the level of selfishness between pairs with strong and weak ties, respectively, while the value of λ_0 corresponds to the percentage of pairs that have strong (or weak) ties.

Policy C *Limit - Rates*: Each pair of nodes exchanges messages with probability p_1 if they contact with rate lower than λ_0, and adjust the message exchange probability if they contact with higher rate. Hence, for all pairs $\{i, j\}$ with $\lambda_{ij} > \lambda_0$, it will hold that $p(\lambda_{ij}) \cdot \lambda_{ij} = p_2 \cdot \lambda_0 = const.$.

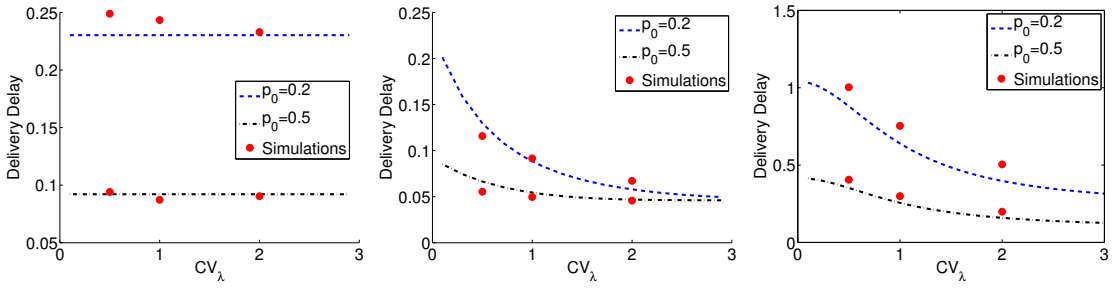
Policy D *Exponential*: Each pair of nodes $\{i, j\}$ exchanges messages with probability $p_0 \cdot (1 - e^{-m \cdot \lambda_{ij}})$, where λ_{ij} is their meeting rate and $p_0 < 1$ and m are positive constants. The message exchange probability is higher for node pairs that meet more frequently.

Table 3.3: Selfishness policies.

Policy A	$p(\lambda) = p_0$	
Policy B	$p(\lambda) = \begin{cases} p_1 & : \lambda \leq \lambda_0 \\ p_2 & : \lambda > \lambda_0 \end{cases}$	$\overline{F}_\lambda(\lambda_0) = p_0$
Policy C	$p(\lambda) = \begin{cases} p_1 & : \lambda \leq \lambda_0 \\ p_2 \cdot \dfrac{\lambda_0}{\lambda} & : \lambda > \lambda_0 \end{cases}$	$\overline{F}_\lambda(\lambda_0) = p_0$
Policy D	$p(\lambda) = p_0 \cdot (1 - e^{-m \cdot \lambda})$	

To find the expected message delivery delay, for a certain network size and a certain routing protocol, only the computation of the effective contact rates' mean value ($\mu_\lambda^{eff.}$) is needed (Result 5). In Table 3.2, we present the closed form expressions for the $\mu_\lambda^{eff.}$ for these selfishness policies, under different mobility patterns. Specifically, we considered three cases for the contact rates distribution f_λ: (i) Gamma, (ii) Exponential[4], and (iii) Pareto distribution. We chose

[4]The Exponential distribution can be defined also as a Gamma distribution with parameters $\alpha = 1$ and $\beta = \mu_\lambda^{-1}$. However, for clarity, we present the results separately.



(a) Policy A (b) Policy B ($p_1 = 0, p_2 = 1$) (c) Policy D ($m = 0.25$)

Figure 3.1: *Delivery Delay* in networks with $N = 100$ nodes and varying *Mobility* characteristics ($\mu_\lambda = 1$ and $CV_\lambda \in [0, 3]$) for three different selfishness policies and under epidemic routing. The theoretical values of Delivery Delay for two parameters (p_0) for each selfishness policy are denoted with dashed lines and the corresponding simulations' average delivery delays are denoted with dots.

to analyze these distributions, because they capture a large range of contact variabilities, and (especially Gamma) were shown to match well the observed contact rates distributions in real social networks [108].

Similar closed form expressions of $\mu_\lambda^{eff.}$, which depend only on the selfishness policy's parameters, $p(\lambda)$, and the first moments of the contact rates distributions, f_λ, can be found as well for other cases of $p(\lambda)$ and f_λ.

3.3.3 Validation

The results derived so far provide us with closed-form predictions for the performance of various protocols and selfishness behaviors under a broad class of mobility models. In Section 3.3.3.1, we first validate their accuracy in (synthetic) scenarios belonging to this mobility class, in order to isolate the effects of the various analytical approximations we have performed towards obtaining the expressions for these otherwise very complex problems. Then, in Section 3.3.3.2 we further consider trace-driven scenarios, where in addition to approximation errors, departures from many, if not most, of the model assumptions are expected to introduce further inaccuracies.

3.3.3.1 Synthetic Simulations

We developed a simulator that generates synthetic networks with mobility conforming to the mobility class of Def. 3: In each scenario, we assign to each pair $\{i, j\}$ a contact rate λ_{ij}, which we draw randomly from f_λ[5] and create a sequence of contact events (according to a Poisson process with rate λ_{ij}). We also assign to $\{i, j\}$ a probability p_{ij} according to the function $p(\lambda)$. Then, we simulate a large number of message exchanges, by choosing randomly for each message the source-destination pair, and calculate the mean simulated delivery delay by averaging the results.

In Fig. 3.1 we present, for networks with $N = 100$ nodes, how the mobility heterogeneity (i.e. $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$) affects the message delivery delay, under different selfishness policies.

[5]In the results we present, the contact rates are drawn from a Gamma distribution, $f_\lambda \sim Gamma$, with variable parameters μ_λ and CV_λ (see Fig. 3.1).

The theoretical results (dashed lines) show (Fig. 3.1(a)) that in the case of uniform selfishness (Policy A), the mobility heterogeneity (i.e. CV_λ) level does not affect the message delivery delay. For the same parameters of $p(\lambda)$ (i.e. p_0), the expected message delivery delay is equal for different mobility heterogeneity scenarios. However, for the non-uniform selfishness policies (Policies B and D), where the selfishness depends on the pairs' contact rates, mobility heterogeneity highly affects the message delivery delay (Fig. 3.1(b) and Fig. 3.1(c)). For the same parameters of $p(\lambda)$, the expected delivery delay decreases as the mobility heterogeneity level increases.

In all cases presented in Fig. 3.1, the synthetic simulations results (dots), are very close to our theoretical predictions, despite the various assumptions and approximations we used in our theoretical analysis. We have also performed simulations for larger networks (i.e. 300 and 1000 nodes), with similar findings.

3.3.3.2 Real-world Traces

In this section, we conduct simulations on the following sets of real mobility traces[6]:
Cabspotting [118]: GPS coordinates from 536 taxi cabs collected over 30 days in San Francisco.
Infocom [125]: Bluetooth sightings of 98 mobile and static nodes (*iMotes*) collected during Infocom 2006.
Sigcomm [115]: Bluetooth sightings of 76 mobile users of the MobiClique application at Sigcomm 2009.

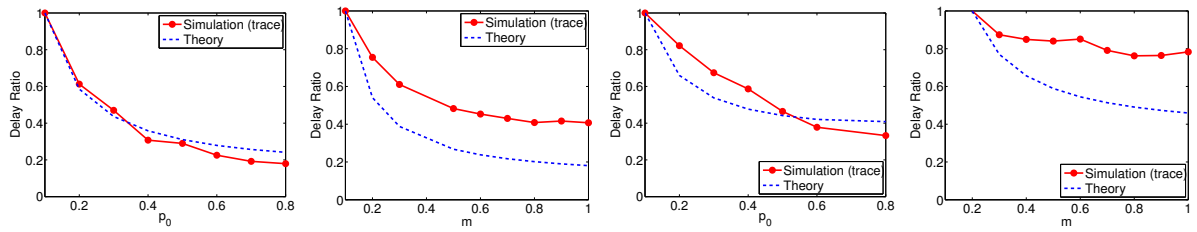
In Fig. 3.2 we show, for the *Cabspotting* and the *Infocom* traces, how the delivery delay of SnW routing decreases as the cooperation between nodes increases. Specifically, we present the relative delay decrease[7], $\frac{E[T_D]}{E[T_D^{max}]}$, i.e. the ratio of the average delivery delay in each scenario ($E[T_D]$) over the delay of the scenario with the highest level of selfishness ($E[T_D^{max}]$).

In Fig. 3.2(a) we simulated scenarios where nodes apply a Policy B selfishness (Table 3.3) with parameters $p_1 = 0, p_2 = 1$ (i.e. only "strong" ties). In each scenario different values of p_0 (i.e. percentage of pairs that cooperate) are selected; higher values of p_0 correspond to scenarios with less selfishness. Results of scenarios where nodes apply a Policy D selfishness are presented in Fig. 3.2(b). It can be seen that for Policy B, the accuracy is significant, while for Policy D, the average simulated delivery delay (red line) decreases slower than predicted (dashed blue line). However, for both policies, the simulation results and theoretical predictions agree qualitatively, even if not always quantitatively.

In the *Infocom* trace (Fig. 3.2(c) and 3.2(d)), the theoretical predictions are less accurate than in *Cabspotting*. The main reason for this, is that the mobility patterns of the Infocom trace deviate from the assumptions of our mobility model more than the mobility patterns of the Cabspotting trace. In particular, we observed higher community structure and temporal characteristics that cannot be captured by a Poisson contact process (i.e. during night, there are almost no contacts).

[6]For brevity, we present here results only on the first two traces and we test our predictions on the Sigcomm trace in following sections.

[7]We present relative values in order to allow a direct comparison between the two traces, whose characteristics (network size, mobility statistics, etc.) differ significantly.



(a) Cabspotting, Policy B (b) Cabspotting, Policy D (c) Infocom, Policy B (d) Infocom, Policy D

Figure 3.2: Relative decrease of delay, $\frac{E[T_D]}{E[T_D^{max}]}$, of SnW routing, in scenarios on the *Cabspotting trace* with (a) Policy B ($p_1 = 0, p_2 = 1$ and variable p_0) selfishness and $L = 10$ copies, and (b) Policy D ($p_0 = 0.2$ and variable m) selfishness and $L = 20$ copies; *Infocom trace* with(c) Policy B ($p_1 = 0, p_2 = 1$ and variable p_0) selfishness and $L = 20$ copies, and (d) Policy D ($p_0 = 0.2$ and variable m) selfishness and $L = 20$ copies.

3.4 Performance and Power Consumption Trade-offs

We have so far considered the effect of different selfishness policies on performance, assuming that the actual policy is *given* (e.g. user preferences, security or privacy concerns, etc.). However, it might be the case that a node's reluctance to always relay 3^{rd} party traffic stems from resource-related concerns (e.g. spending energy). In this case, the selfishness policy could be seen as a way for the node to control the amount of resources (e.g. transmission power) contributed to participate in the network.

Moreover, nodes would not object to use a different policy, e.g. one that improves the network-wide, and thus average node performance, if it would not result in a higher expected resource consumption for them. For instance, if with a policy x and a policy y, a node consumes the same energy, but the message delivery probability achieved by policy x is higher than this of y, i.e. $P_x > P_y$, then it could choose to apply policy x in order to improve the overall network performance.

To this end, in this section, we examine the extent to which nodes could achieve different tradeoffs between resource consumption and network performance, using different policies, in two generic communication scenarios. At first, using a simple communication traffic injection model, we investigate the tradeoff between *Delivery Delay* and *Power Consumption* (Section 3.4.1). In the second case, we turn our attention to the possible *Delivery Probability - Power Consumption* tradeoffs that can be achieved by an opportunistic content sharing mechanism (Section 3.4.2).

3.4.1 Delivery Delay vs Power Consumption

As mobile devices rely on their batteries, whose energy capacity is limited, power consumption becomes a crucial issue. Nodes might prefer saving energy resources than consuming a significant amount of them for network operations (i.e. storing and relaying messages).

Nevertheless, the total power consumption for relay traffic does not only depend on the policy choice, but also on the total message load in the network, and the protocol used. In order for a node to be able to estimate the expected power overhead of a given policy, we need to "level the ground", in a sense, and define a simple traffic model (see Table 3.4 for notation) that will allow us to compare directly the power overhead of different policies.

Let us assume that there are (on average) N_f number of flows in the network, i.e. N_f number of source-destination pairs that exchange messages (we assume that sources are "backlogged",

Table 3.4: Notation for the Communication Traffic Model

TTL	Message lifetime
N_f	Nb of flows
M	Window length (in nb of messages)
$E[N_t^{msg}]$	Avg nb of message transmissions per message and per node
T_∞	Observation time
N_m	Nb of generated messages during time interval $[0, T_\infty]$
$E[N_t]$	Avg nb of transmissions per node in the time interval $[0, T_\infty]$
E_t	Avg energy consumption of a message transmission
P	Avg power consumption

i.e. always have messages to transmit). In order to ensure that source nodes do not insert new messages (input rate) faster than the network can deliver (output rate), some flow control mechanism is needed.

Some works suggest the use of an "out-of-band" channel (e.g. cellular network) for acknowledgements [6]. In this case, each source node could be forced, e.g. to not send a new message before the previous message is ACKed. In fact, we could also assume a window of M messages per flow that can go unacknowledged before a new message is send. Thus, if $E[T_D]$ is the expected delay of a message, the total load per flow is M messages per $E[T_D]$ time units (assuming an instant acknowledgement). If on the other hand, a slower "in-band" flow control is used, the RTT could also be expressed as $c \cdot E[T_D]$, $c > 1$.

Alternatively, each message can be assigned a *message lifetime* value, i.e. a TTL, after which the message cannot be forwarded or delivered to the destination, and nodes can drop any copies of it, in order to release valuable storage space[8]. To achieve a high message delivery probability (i.e. $PDR \approx 1$), the message lifetime must be set as

$$TTL = c_{TTL} \cdot E[T_D] \tag{3.9}$$

and c_{TTL} is large enough, such that the probability that the TTL expires before the message is delivered to the destination is small (this is necessary since we are interested in this section on the message delay).

Regardless of the exact flow control policy used (not of interest to this work), the above discussion suggests that a reasonable model for (stable) traffic loads is to assume that each source injects on average M new packets for each time interval $c \cdot E[T_D]$ (with c dependent on the flow control policy). In following, without loss of generality, we will assume a TTL flow control mechanism.

Under the condition of large $c = c_{TTL}$ (i.e. value of TTL such as $PDR \approx 1$), we can easily show that the average number of transmissions per message a node has to perform, $E[N_t^{msg}]$, depends only on the routing protocol, $\mathcal{P}$, and the network size, N, and is independent of the message delivery delay, i.e.:

$$E[N_t^{msg}] = c_t \tag{3.10}$$

[8]The lack or volatility of end-to-end paths in opportunistic networks, implies that the implementation of a transport protocol with feedback per packet (e.g. as ACK messages in TCP), as described above, might be either inefficient or infeasible. As a result, the TTL can often be used as an implicit flow control, allowing up to M new packets per TTL for each flow.

where c_t is a constant dependent on N and $\mathcal{P}$. As an example, in *Spray and Wait* routing with L copies, as there are in total approximately L messages transmissions ($L-1$ to relay nodes and 1 to destination node) before the expiry of the TTL, the average number of message transmissions per message and per node is given by $E[N_t^{msg}] = \frac{L}{N}$.

Hence, if we observe our network with communication traffic as described above, for a long time period T_∞, it follows easily that the expected number of generated messages is

$$N_m = N_f \cdot M \cdot \frac{T_\infty}{c \cdot E[T_D]}. \tag{3.11}$$

As the number of transmissions per generated message a nodes does is $E[N_t^{msg}]$, the total number of transmissions a node does in the time interval $[0, T_\infty]$ is

$$E[N_t] = N_m \cdot E[N_t^{msg}] = N_f \cdot M \cdot \frac{T_\infty}{c \cdot E[T_D]} \cdot c_t \tag{3.12}$$

where we substituted from Eq. (3.10) and Eq. (3.11)

Then, the power consumption rate can be calculated as

$$P = \frac{\text{Total Energy Consumption in } [0, T_\infty]}{T_\infty} = \frac{E_t \cdot E[N_t]}{T_\infty} \tag{3.13}$$

where E_t is the average energy for a single message transmission. The following result follows after substituting Eq. (3.12) in Eq. (3.13).

Result 6. *The average node power consumption is inversely proportional to the average message delivery delay and is given by*

$$P = c_p \cdot \frac{1}{E[T_D]} \tag{3.14}$$

where $c_p = \frac{E_t \cdot N_f \cdot M \cdot c_t}{c}$.

In Result 6, c_p is a constant that depends on the (i) network size N, (ii) the protocol used $\mathcal{P}$, (iii) the message size (E_t) and (iv) the traffic intensity (N_f, M). However, c_p is independent of the *selfishness policy* and the *mobility* of the nodes. Therefore, the main implication that comes of Result 6, is that:

Corollary 2. *In a Heterogeneous Contact Network, no matter how simple or sophisticated the selfishness policy used, the achievable power-delay operating regimes are exactly the same; In other words, whatever power-delay tradeoff can be achieved by some socially selfish policy, can also be achieved by the simple uniform policy.*

The above conclusion is somewhat surprising at first, given the range of strategies available under our social selfishness definition. However, we will try to shed some light on this counterintuitive result: Let assume a relay node i with some messages in its buffer. At the next contact event, i will forward each of the messages, e.g. to node j, with some probability, which depends on the protocol and the state of j (i.e. if j has the message or is the destination, etc.)[9]. It, then, follows that the more (effective) contact events a node has, the more messages it will

[9]Similarly, i will receive a message from j with some probability. Since we assume backlogged sources, the number of messages in the buffer of each node will be *on average* the same.

transmit (i.e. power consumption). Since all nodes apply the same policy, the average number of contact events per time unit (and thus the power consumption) is the same for every node and $\propto E[p(\lambda) \cdot \lambda] \equiv \mu_{\lambda}^{eff.}$. Now, considering the discussion and results in Section 3.3, which show that the delivery delay is inversely proportional to $\mu_{\lambda}^{eff.}$, the relation suggested by Result 6 becomes evident.

To this direction, we can derive the following result (by simply combining Results 5 and 6) that relates the power consumption with the selfishness policy and mobility characteristics:

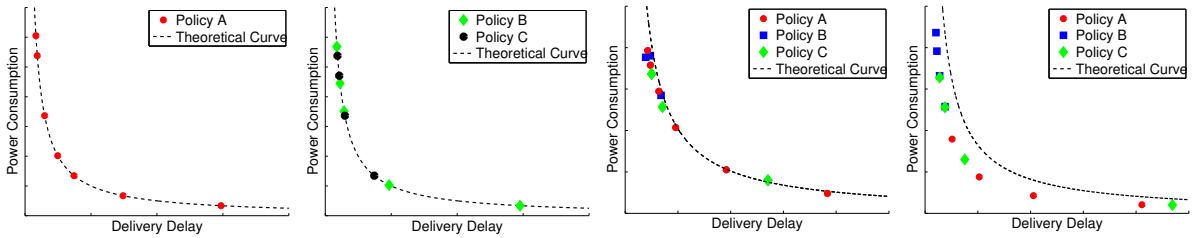
Result 7. *The average node power consumption in an Heterogeneous Contact Network is approximately given by*

$$P = \frac{c_p}{c(N, L)} \cdot \mu_{\lambda}^{eff.} \tag{3.15}$$

where $c(N, L)$ and c_p are defined in Results 5 and 6, respectively.

Thus, the expressions in Table 3.2 can be used to compute the average node power consumption, under the selfishness policies of Table 3.3.

3.4.1.1 Validation



(a) Synthetic: uniform policy (b) Synthetic: non - uniform policies (c) Infocom (d) Sigcomm

Figure 3.3: Power consumption - message delivery delay trade off. Synthetic simulations with (a) uniform and (b) non-uniform selfishness policies. Simulations on the (c) Infocom and (d) Sigcomm real traces of both uniform and non-uniform selfishness policies scenarios.

From Result 6 we can see that the relation between power consumption and message delivery delay can be described by a reciprocal function or by a curve of the form $y = \frac{a}{x}$.

To investigate how accurate this prediction is, we first consider a heterogeneous mobility scenario ($f_{\lambda} \sim Gamma$, $\mu_{\lambda} = 1$, $CV_{\lambda} = 1$), consisting of 100 nodes. We generate communication traffic between node pairs, according to the rules of the traffic model described in Section 3.4.1, and select SnW with $L = 10$ copies as the routing protocol. We perform Monte Carlo simulations. At first, we simulate scenarios with the uniform selfishness policy (Policy A) and choose values for the selfishness intensity (i.e. p_0) spanning the range $(0, 1)$, i.e. for minimum to maximum power consumption. Fig. 3.3(a) shows the simulation results for some sample values of p_0. It can be seen there that these exactly match our theoretical predictions.

We then simulate scenarios with different, non-uniform selfishness policies, in order to examine whether the delay-power curve is indeed the same or not. As is evident by Fig. 3.3(b), the simulated results for both non-uniform policies considered also coincide with the theoretical

curve, which is also the delay-power curve for the uniform policy. In other words, by changing selfishness policies and their parameters, one can only achieve a shift on the theoretical curve.

To further examine the validity of this interesting finding, we test our predictions also in two real-world scenarios, the *Infocom* and *Sigcomm* traces. In Fig. 3.3(c), we use SnW routing with $L = 5$ copies in the Infocom trace and we create traffic conditions as described earlier. We measured the delivery delay of the messages and the power consumption of the nodes and plot the achievable delay-power tradeoff points for different policies. As it can be seen, our qualitative finding also holds here (i.e. all policies seem to have the same achievable region), and experimental values are quite close to the theoretically predicted curve. Similar observations can be made for the results of simulations on the Sigcomm trace (Fig. 3.3(d)). In this trace, although the theoretical curve seems to be a slightly displaced, it is clear that all policies also lie on the same tradeoff curve, as predicted.

3.4.2 Delivery Probability vs Power Consumption

In the previous section, we showed that the region of possible tradeoffs between *Delivery Delay* and *Power Consumption* is not affected by the selfishness policy. A key question arising then is: *is there not a way to achieve better performance-power tradeoff regions, e.g. compared to the uniform policy, by intelligently choosing the selfishness policy?*

In order to further explore this question, we turn our attention to another metric of high importance, namely the *delivery probability* of a message (or Probability Delivery Ratio, *PDR*). Thus, in this section, we investigate the *PDR - Power Consumption* tradeoffs using another example application, namely *content sharing* in opportunistic networks. The rationale behind this choice is twofold: first because content-centric applications have attracted increasing attention in both wired and wireless networks, and second to demonstrate the applicability of our framework to non end-to-end communication scenarios.

3.4.2.1 Opportunistic Content Sharing

In content sharing scenarios, new messages might be useful only for some fixed amount of time (e.g. related to the content nature), and interested nodes would like to *access* such messages before this time. We assume that there are $N_A (\leq N)$ nodes, in the network, that hold a content A for which another node i is interested in. This content can be data (e.g. a map, news, video, etc.) or even a service that these nodes can provide (e.g. Internet access or a computing service [26]). We also assume that this content can be only delivered directly when node i contacts any of the N_A nodes with the content, and not through relay nodes (this assumption might related to protocol complexity, but often comes very natural, as for example, when the content is an actual computing service the N_A providers can offer)[10].

The following result, gives the probability for a node i to successfully access content A by some time T.

Result 8. *In a Heterogeneous Contact Network with selfishness policy $p(\lambda)$, if N_A nodes hold a content A, then the probability for another node to access the content by a time T, is given by*

$$P_A\{T\} = 1 - \left(E\left[e^{-\lambda \cdot p(\lambda) \cdot T}\right]\right)^{N_A} \tag{3.16}$$

[10]Note that the selfishness policy applies even in this direct case, since e.g. content providers might not be equally willing to service or forward to any interested node.

where the expectation is taken over f_λ.

Proof. Let us denote $P_a\{j,T\}$ the probability the node i to contact a node $j \in [1,...,N_A]$ and exchange messages with it (i.e. effective contact event) before a certain time T. Obviously $P_a\{j,T\}$ (i) depends on the contact rate λ_{ij} and the selfishness policy $p(\lambda)$, and (ii) as the inter-contact intervals are exponentially distributed it is given by[11]

$$P_a\{j,T|\lambda_{ij},p(\lambda_{ij})\} = 1 - e^{-\lambda_{ij}\cdot p(\lambda_{ij})\cdot T}$$

Since the probability a node to have the content is the same for all nodes, we can write

$$P_a\{j,T\} = \int_0^\infty P_a\{j,T|\lambda_{ij},p(\lambda_{ij})\}\cdot f_\lambda(x)dx = \int_0^\infty \left(1 - e^{-\lambda_{ij}\cdot p(\lambda_{ij})\cdot T}\right)\cdot f_\lambda(x)dx \;\; = \;\; 1 - E\left[e^{-\lambda\cdot p(\lambda)\cdot T}\right]$$

where the expectation in is taken over f_λ.

Node i will not access the content by time T, only if it does not contact any of the N_A nodes. Hence, we can write for the probability that i will get the content by time T:

$$P_A\{T\} \;=\; 1 - \overline{P}_A\{T\} \;=\; 1 - \prod_{j=1}^{N_A}\overline{P}_a\{j,T\} \;\; = \;\; 1 - \prod_{j=1}^{N_A}\left(1 - P_a\{j,T\}\right)$$

where $\overline{P}$ denotes the probability of the complementary event. Now, combining the above two equations and the fact that the nodes j with the content (and the respective contact rates λ_{ij}) are independent, it follows

$$P_A\{T\} = 1 - \prod_{j=1}^{N_A}\left(1 - \left(1 - E\left[e^{-\lambda\cdot p(\lambda)\cdot T}\right]\right)\right) = 1 - \prod_{j=1}^{N_A} E\left[e^{-\lambda\cdot p(\lambda)\cdot T}\right] \;\; = \;\; 1 - \left(E\left[e^{-\lambda\cdot p(\lambda)\cdot T}\right]\right)^{N_A}$$

□

Closed form expressions for the probability $P_A\{T\}$ under different selfishness policies (Table 3.3) and mobility patterns (f_λ) can be found in Table 3.5.

We know from Result 7 that the average power consumption is proportional to $\mu_\lambda^{eff.}$. However, the expression for the content delivery probability (Result 8) relates to the mobility pattern and the selfishness policy in a non-linear way, that is also more complex than the case of delay. The first observation is that it's not easy to deduce a simple relation between the power consumption and the PDR, under generic mobility and selfishness characteristics, as was the case for power and delay (Result 6). The non-linearity also implies that it might now be possible indeed to change (and ultimately improve) the achievable power - performance (PDR) region.

[11] The CDF of an exponential distribution with rate λ is given by $F(x) = 1 - e^{-\lambda\cdot x}$.

Table 3.5: Probability a node to access the content by time T, $P_A\{T\}$.

Policy A	Policy B	Policy C
Gamma: $f_\lambda(x) = \frac{\beta^a}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta x}, \quad \alpha, \beta, x > 0$ $\left(\text{where} \quad \alpha = \frac{1}{CV_\lambda^2}, \quad \beta = \frac{1}{\mu_\lambda \cdot CV_\lambda^2}\right)$		
$1 - \left(1 + p_0 \cdot T \cdot \mu_\lambda \cdot CV_\lambda^2\right)^{-\frac{N_A}{CV_\lambda^2}}$	$1 - \left(\frac{\gamma\left(\frac{1}{CV_\lambda^2}, \left(p_1 \cdot T + \frac{1}{\mu_\lambda CV_\lambda^2}\right) \cdot \lambda_0\right)}{\Gamma(\frac{1}{CV_\lambda^2}) \cdot \left(1 + p_1 \cdot T \cdot \mu_\lambda \cdot CV_\lambda^2\right)^{\frac{1}{CV_\lambda^2}}} + \frac{\Gamma(\frac{1}{CV_\lambda^2}) - \gamma\left(\frac{1}{CV_\lambda^2}, \left(p_2 \cdot T + \frac{1}{\mu_\lambda CV_\lambda^2}\right) \cdot \lambda_0\right)}{\Gamma(\frac{1}{CV_\lambda^2}) \cdot \left(1 + p_2 \cdot T \cdot \mu_\lambda \cdot CV_\lambda^2\right)^{\frac{1}{CV_\lambda^2}}} \right)^{N_A}$	$1 - \left(\frac{\gamma\left(\frac{1}{CV_\lambda^2}, \left(p_1 \cdot T + \frac{1}{\mu_\lambda CV_\lambda^2}\right) \cdot \lambda_0\right)}{\Gamma(\frac{1}{CV_\lambda^2}) \cdot \left(1 + p_1 \cdot T \cdot \mu_\lambda \cdot CV_\lambda^2\right)^{\frac{1}{CV_\lambda^2}}} + e^{-p_2 \lambda_0 \cdot T} \cdot p_0 \right)^{N_A}$
	where $\gamma\left(\frac{1}{CV_\lambda^2}, \frac{\lambda_0}{\mu_\lambda CV_\lambda^2}\right) = \Gamma\left(\frac{1}{CV_\lambda^2}\right) \cdot (1 - p_0)$	
Exponential: $f_\lambda(x) = \frac{1}{\mu_\lambda} \cdot e^{-\frac{x}{\mu_\lambda}}, \quad \mu_\lambda, x > 0$		
$1 - (1 + p_0 \cdot T \cdot \mu_\lambda)^{-N_A}$	$1 - \left(\frac{1 - p_0^{1 + p_1 \cdot T \cdot \mu_\lambda}}{1 + p_1 \cdot T \cdot \mu_\lambda} + \frac{p_0^{1 + p_2 \cdot T \cdot \mu_\lambda}}{1 + p_2 \cdot T \cdot \mu_\lambda} \right)^{N_A}$	$1 - \left(\frac{1 - p_0^{1 + p_1 \cdot T \cdot \mu_\lambda}}{1 + p_1 \cdot T \cdot \mu_\lambda} + p_0^{1 + p_2 \cdot T \cdot \mu_\lambda} \right)^{N_A}$

3.4.2.2 Evaluation

To obtain some useful evidence, we will focus here on two selfishness policies, namely the *uniform* policy (Policy A) and the *limit-rates* policy (Policy C). Our choice for the specific non-uniform selfishness policy is based on the fact that it was proposed in [64] as a policy designed for a content dissemination application, which resembles the application we study.

From its definition (Table 3.3), we can see that Policy C limits the average number of effective contacts for pairs that contact more frequently than a certain threshold. The intuition behind this mechanism, is that a node i avoids communicating every time with the nodes j with whom it meets frequently, because (i) each effective contact incurs some energy consumption, and (ii) as they meet frequently, the probability node j to hold a content message in which node i is interested in and which did not exist in the memory of j their previous contact event, is small. Thus, limiting the effective contact events with frequently met nodes would result in a better PDR-power tradeoff.

Our theoretical predictions among with simulated results from two scenarios where we assign a content to random nodes and measured the delivery probability of it to a certain node, are presented in Fig. 3.4. These confirm the intuition about the superiority of Policy C regarding content sharing applications. In Fig. 3.4(a) we present the PDR values in scenarios with uniform and rate-limit selfishness policies, where only one node holds the content message. As it can be seen, Policy C achieves always higher PDR than policy A for the same power consumption values. Specifically, Fig. 3.4(b) shows the improvement (i.e. the ratio $\frac{PDR_C - PDR_A}{PDR_A}$) in PDR we achieve with Policy C, which, for some values of power consumption, is almost 30%. In some other scenarios we simulated, this improvement was even up to 70%.

Fig. 3.4(c) and 3.4(d) present the comparison of the two policies, in a scenario with more heterogeneous mobility ($CV_\lambda = 2$) where $M = 5$ nodes hold the content. The observations about the performance of the two policies remain the same.

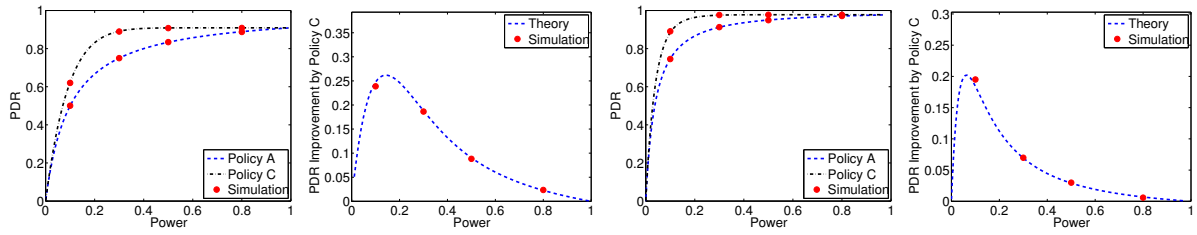
Finally, it is evident in Fig. 3.4 that simulations results for the synthetic heterogeneous model (red dots) match our theoretical predictions very well.

As a final step, we test again the accuracy of our findings in two real-networks, the *Sigcomm* and *Infocom* traces. We simulated scenarios with different number of content holders and for the same selfishness policies as before. The results are presented in Fig. 3.5 and compared to the theoretical prediction. As it can be observed, while the absolute values do not match exactly, Policy C again outperforms the uniform policy, and the relative performance improvement follow the shape of the theoretical curve quite well (this is very important when considering finding optimal operating points, using the theoretical curve).

Hence, we can conclude that our model can provide quite accurate predictions, even for real network scenarios. Finally, it is clear that, unlike the case of delay-power tradeoff, using social selfishness wisely can improve performance here, and our model could be used in order to predict the relative performance of different policies and, consequently, for policy optimization.

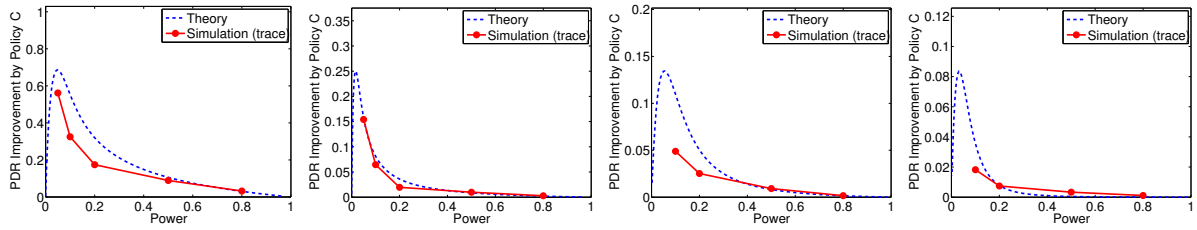
3.5 Related Work

The feasibility of communication over a MSN highly depends on the willingness of nodes to cooperate. To this end, many techniques and protocols were proposed in order to motivate nodes to act as relays for messages that are not generated by or destined to them [22, 74, 94, 127, 145]. In [94] a reputation mechanism is used to encourage nodes to cooperate in order (i) to be able to



(a) $M = 1$, $T = \frac{10}{\mu_\lambda}$ (b) $M = 1$, $T = \frac{10}{\mu_\lambda}$ (c) $M = 5$, $T = \frac{5}{\mu_\lambda}$ (d) $M = 5$, $T = \frac{5}{\mu_\lambda}$

Figure 3.4: (a),(c) Probability Delivery Ratio of a content of policy A selfishness (blue) and policy C selfishness (black) for different power consumption levels. (b),(d) Relative difference of the Probability Delivery Ratio between Policy C and Policy A selfishness, i.e. $\frac{PDR_C - PDR_A}{PDR_A}$. Mobility characteristics: $\mu_\lambda = 1$; (a),(b) $CV_\lambda = 1$ and (c),(d) $CV_\lambda = 2$.



(a) Sigcomm, $M = 3$, $T = \frac{20}{\mu_\lambda}$ (b) Sigcomm, $M = 5$, $T = \frac{20}{\mu_\lambda}$ (c) Infocom, $M = 3$, $T = \frac{10}{\mu_\lambda}$ (d) Infocom, $M = 5$, $T = \frac{10}{\mu_\lambda}$

Figure 3.5: Relative difference of the Probability Delivery Ratio between Policy C and Policy A selfishness, i.e. $\frac{PDR_C - PDR_A}{PDR_A}$, in the *Sigcomm trace* with (a) $M = 3$, (b) $M = 5$ number of copies and $T = 20/\mu_\lambda$, and in *Infocom trace* with (c) $M = 3$, (d) $M = 5$ number of copies and $T = 10/\mu_\lambda$.

receive the messages destined to them, and (ii) the other nodes to offer them their services (i.e. relay their messages). Another approach, which results in growing incentives to nodes for acting as relays, is followed in [127], where each node i is willing to forward the messages of another node j, according to the number of messages node j has forwarded already for it. Finally, credit-based mechanisms are presented in [145] and [22], as well as barter-based mechanisms in [74].

Furthermore, many analytical and simulation-based studies investigate the effects of node selfishness on communication performance [61, 69, 86, 88, 106]. In [106] authors investigate, through simulations, how the performance of epidemic schemes is affected when the network comprises of non-cooperative nodes. They consider two kinds of selfishness: in each contact event either nodes are unwilling to copy a message with probability p_{nc} or they are unwilling to forward it with probability p_{nf}. For a similar scenario, Karaliopoulos [69] models probabilistic selfish behavior of nodes in homogeneous networks (i.e. constant contact rate λ for every pair of nodes). In [86] authors extend the work of [69] in terms of multicast applications. The authors of [88] model selfishness in a different scheme, where each node transmits its own message (operates as a source) with probability p and transmits one of the messages it has as a relay with probability 1-p. They assume that only one message can be exchanged per contact event and use only 2-hop routing. Another approach of selfishness is tackled in [61], where authors propose a selfishness model where each node is either selfish or altruist (modeled as a probability p_i for each node i) regarding all its contacts (i.e. i shows the same selfishness for every other node j, $p_{ij} = p_i$) and investigate through simulations the effect on the communication throughput.

The above protocols and studies, assume (under different models) that every node is either totally selfish or not. However, the assumption that users are selfish and are not willing to forward packets for anyone else, might not always hold. In this direction, the notion of *social selfishness* appears. In social selfishness, the nodes might be selfish only regarding some other nodes with which they have a weak (or even a strong) social relation ("tie") [64, 82, 84]. In [84] authors use a model of a network with two communities and introduce the notion of selfishness that depends on the contact rate between nodes. For nodes with high contact rate (e.g. within the same community) the selfishness is characterised by the probability p_i and for nodes with low contact rate another value p_o for selfishness is considered. They build a Markov Chain and investigate the effect on the performance through simulations. In [64] the authors investigate the role of the "weak ties" (i.e. pairs of nodes that contact infrequently, which in our case means the pairs of nodes with small contact rate λ_{ij}) in a content updating/dissemination scenario. Finally, in [82] a routing protocol, designed for networks where nodes have social selfishness behaviors, is proposed.

Our work, being the first to provide a theoretical framework and analytical closed-form results, complements previous studies on the effect of social selfishness on communication performance, which are limited to evaluation through simulations [64, 82] or analytical modeling of specific cases [84]. Moreover, not only the heterogeneous mobility model we consider can capture much wider range of scenarios than the models used in previous analytical studies [69, 84, 88], but also our results were shown to capture (either qualitatively or quantitatively) the much more complex characteristics of real-networks' mobility.

3.6 Conclusion

In this chapter, we analysed the effect of social selfishness on opportunistic communications. Based on the model for heterogeneous mobility presented in Chapter 2, we built a generic model that can describe a wide range of common social selfishness behaviors (related to privacy concerns, resources consumption, etc.). Based on our mobility / selfishness framework, we derived closed form results for predicting the message delivery delay in a network with (socially) selfish nodes. Furthermore, we investigated how selfishness affects the performance - power consumption tradeoffs in a network, under two communication scenarios. We derived results that show if and when it is possible to optimize a selfishness policy in order to achieve better tradeoffs.

Due to the lack of existing solutions fighting social selfishness, we deem as essential to have an analytical framework for it and predict the performance degradation it causes on message dissemination, which as shown depends on various factors (selfishness behaviors and mobility). We believe that our work can be a useful tool for the design of novel protocols and applications for socially selfish environments.

Chapter 4

Modeling and Analysis of Communication Traffic Heterogeneity in MSNs

4.1 Introduction

As discussed earlier, node mobility plays a major role both in the performance and the design of protocols and applications for MSNs, and a lot of effort has been made recently to capture and model the heterogeneous mobility patterns of real networks [76, 109, 112, 113]. On the contrary, the communication traffic patterns used in studies of MSNs have not received an equal amount of attention.

It is usually assumed, implicitly or explicitly, that all traffic is uniform: each pair of nodes exchanges the same amount of messages. However, intuition suggests that traffic between nodes, just like mobility, cannot be expected to be homogeneous either. This is also supported by empirical studies on social networks [54, 138], where the frequency of message exchanges might widely vary among pairs of nodes. Further, nodes that have a social relation or reside/move in the same areas, often tend to exchange more messages than others. Therefore, a number of interesting questions arise:

How should one model the heterogeneity in communication traffic? Do heterogeneous traffic patterns affect the performance of information dissemination mechanisms and to what extent?

Towards answering these questions, in this chapter we investigate *if*, *when* and *how* traffic patterns affect the communication performance in mobile social networks. Specifically:

• We examine what characteristics of traffic heterogeneity can have an effect on performance, and show that only when (end-to-end) traffic demand is correlated with pairwise contact rates performance is affected. Based on these findings, we propose an analytically tractable model that can describe a large range of non-uniform traffic patterns (Section 4.2).

• We derive analytical expressions for calculating the joint effect of traffic and mobility heterogeneity in the performance of basic forwarding mechanisms (Section 4.3).

• We use these expressions to show that the common understanding about these mechanisms, e.g. the gains from having additional replicas, might radically change when traffic is heterogeneous (Sections 4.3.2 and 4.3.3).

• We validate our analytical findings through simulations (Section 4.4.1) and, by apply-

ing them to datasets of real-world networks that contain information about the mobility and communication patterns of participating nodes (Section 4.4.2).

• Finally, we present possible extensions of our study (Section 4.5).

To our best knowledge, this is the first attempt to model end-to-end traffic heterogeneity and analytically study its (quantitative and qualitative) effects on the performance of communications in MSNs. Our analytical findings, as well as simulation results, reveal important aspects of mobile social networking that have not been explored or have not been taken into account in previous studies:

– When frequently meeting node pairs tend to exchange (on average) more/less traffic than other nodes, the communication performance can considerably differ from the homogeneous case. Taking into consideration such traffic patterns allows to better design or tune routing protocols.

– The effects on some forwarding mechanisms, like Direct Transmission [130], can be significant, while at the same time flooding (e.g. Epidemic [137]) or routing (e.g. Spray and Wait [129], EBR [98]) protocols are less affected. In particular, an increasing amount of heterogeneity closes the performance gap between the best (Epidemic) and the worst (Direct Transmission) forwarding.

– Under certain conditions, the impact of traffic heterogeneity can be so important, that it can lead to a reconsideration of the employed communication mechanisms, and even the feasibility of applications (e.g. online social messaging, file sharing, service composition) over a MSN.

4.2 Communication Traffic Model

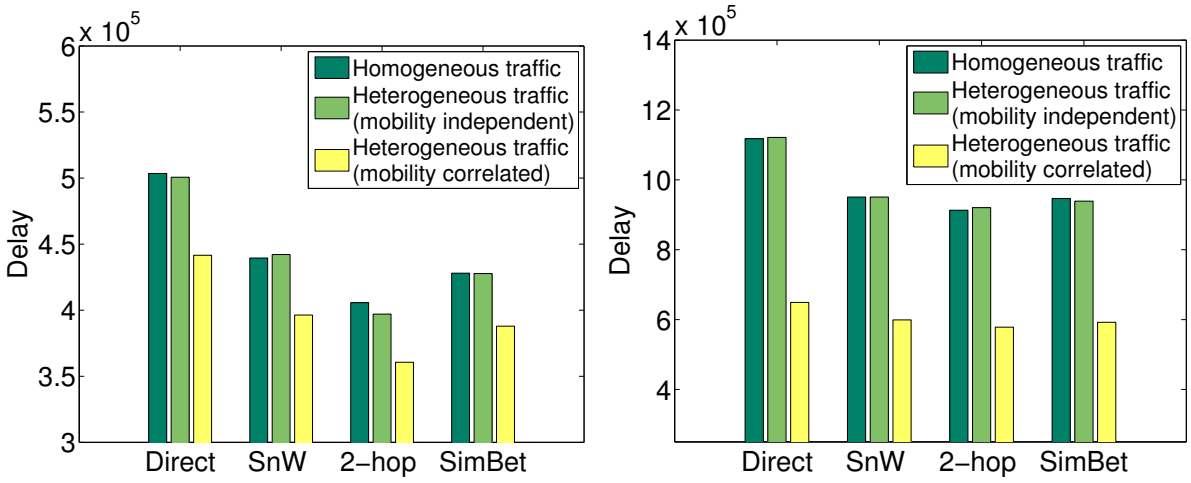
We consider a network $\mathcal{N}$ with N nodes, which communicate in an opportunistic way. Since data exchange is subject to nodes mobility and the resulting *contact events*, we first need to define the mobility model to be used. Similarly to previous chapters, in order to capture mobility heterogeneity and, at the same time, to perform an analytical performance evaluation, we assume nodes to contact each other according to Def. 3.

In addition to who contacts whom and how often, another major question that should be raised when evaluating communication schemes in MSNs (but rarely is) is *who wants to communicate with whom* and *how much traffic do they exchange*?

Intuition suggests that every pair of nodes will not exchange the same amount of traffic. To support intuition, studies from fields related to technological and social networks [37,54,138] have demonstrated the existence of heterogeneous traffic patterns. The same studies further suggest that this heterogeneity depends on the *spatial* and *social* characteristics of these networks. Since *location-based services* [104] and *social networking* [116] are considered among the major applications supported by MSNs, such traffic dependencies on social and/or spatial factors are very probable to appear. What is more, mobility characteristics have also been found to depend on spatial and social characteristics [31, 41, 109]. This clearly seems to argue for a non-homogeneous traffic model. Moreover, traffic and mobility in such networks are expected to exhibit some correlations [54, 138].

Before we proceed to choose a traffic model, one should consider the following questions: *Would the mere heterogeneity of traffic suffice to affect performance? Is it necessary to consider traffic and mobility correlations?*

As stated earlier, information dissemination is determined by the sequence of contact events.



(a) Gowalla dataset

(b) Strathclyde dataset

Figure 4.1: Mean delivery delay of 4 routing protocols, namely *Direct Transmission*, *Spray and Wait* (*SnW*), *2-hop*, and *SimBet*, on the (a) Gowalla and (b) Strathclyde datasets.

Hence, if traffic characteristics are independent of node mobility, one might expect a limited impact on performance.

Towards examining the validity of the above argument, we decided to compare the performance of some well-known opportunistic protocols (direct transmission [130], spray and wait [129], 2-hop routing [43], and SimBet [29]) through simulations on two real traces (we discuss the traces in more detail, later, in Section 4.4), for three traffic scenarios: (i) homogeneous traffic: every pair of nodes has the same chance of being chosen as the source-destination pair for the next message; (ii) heterogeneous traffic that is *mobility independent*: we assign randomly to each pair a different end-to-end traffic demand (with the normalized message generation rate for a pair drawn uniformly in $[1, 1000]$); (iii) heterogeneous traffic that is *mobility dependent*: end-to-end traffic between two nodes is proportional to their contact rate. We generated an equal (sufficiently large) number of messages for all scenarios.

Results for the mean message delivery delay are shown in Fig. 4.1. As is evident from these results, when traffic heterogeneity is independent of mobility (middle bar), the average delay is practically the same to the homogeneous case (left bar), *for all protocols, and across all scenarios* (including additional ones we have tried). In contrast, when traffic is heterogeneous *and* correlated with the contact rates (rightmost bar), Fig. 4.1 shows a clear difference in average delay for all scenarios and protocols. These results provide an initial answer to the above questions:

It is not traffic heterogeneity itself that affects performance, but rather the joint effect of mobility and traffic (heterogeneity).

In other words, unless differences in traffic demand correspond also to differences in contact frequency (e.g. frequently meeting pairs tend to also consistently generate more/less traffic for each other), end-to-end performance will not be affected. This statement is also formally proven in Lemma 5 (Section 4.8.1).

The above observation, together with the initial insight coming from real datasets, motivates us to propose the following simple, yet quite generic, model for end-to-end traffic.

Definition 9 (Heterogeneous Communication Traffic)**.** *The end-to-end traffic demand (per time unit) between a pair of nodes $\{i, j\}$, is a random variable τ_{ij}, such that $E[\tau_{ij}] = \tau(\lambda_{ij})$, where $\tau(\cdot)$ is a continuous function from $\mathbb{R}^+$ to $\mathbb{R}^+$.*

Hence, traffic demand between node pairs can differ and is *on average* correlated with the nodes' contact rate. However, τ_{ij} itself is still random, allowing some node pairs to have little traffic demand even if they meet often (e.g. "familiar strangers"). Furthermore, through the function $\tau(\cdot)$ one can introduce a number of different types and amounts of (positive or negative) correlations between traffic and mobility. While real mobility and traffic patterns are clearly expected to have a number of additional nuances and details, not captured by the models of Def. 3 and Def. 9, respectively, it turns out that these abstractions are still "rich" enough to allow us to draw useful conclusions.

4.3 Analysis

Consider now a MSN with mobility and traffic as defined in the previous section. To calculate a performance metric for this network, e.g. the expected delay, one would consider a large number of messages generated between various source-destination pairs. Therefore, one would further need to know the contact rates between the sources and destinations of these messages. If a message was equally likely to be generated between any pairs of nodes, then the contact rate between the source and destination of this message should be distributed as f_λ (Def. 3). However, if messages are more likely to come from a frequently meeting pair rather than an "average" pair, then the source-destination contact rate (we refer to it as the *effective* contact rate) would be biased towards higher values.

To this end, we derive the following basic proposition for the probability distribution of the *effective* contact rates between source destination node pairs.

Proposition 1. *The probability density function f_τ of the contact rate between the source and the destination $\{s, d\}$ of a random message, in a network following Def. 3 and Def. 9, converges as follows:*

$$f_\tau(x) \xrightarrow{p} \frac{1}{\mathcal{C}} \cdot \tau(x) \cdot f_\lambda(x) \tag{4.1}$$

where $f_\tau(x)dx = P\{\lambda_{sd} \in [x, x+dx)\}$, $\xrightarrow{p}$ denotes convergence in probability, and $\mathcal{C} = E[\tau(\lambda)] = \int_0^\infty \tau(x) f_\lambda(x) dx$ is a normalizing constant.

Proof. Consider a network $\mathcal{N}$ with N nodes. Let $d\lambda = O\left(\frac{1}{N}\right)$, and define the set of nodes with contact rate $\lambda_{ij} \in [\lambda, \lambda + d\lambda)$:

$$\mathcal{N}(\lambda) = \{\{i, j\} : i, j \in \mathcal{N}, \lambda \leq \lambda_{ij} < \lambda + d\lambda\},$$

The total number of messages generated per time unit between pairs $\in \mathcal{N}(\lambda)$ is equal to

$$T(\lambda) = \textstyle\sum_{\{i,j\} \in \mathcal{N}(\lambda)} \tau_{ij} \tag{4.2}$$

where τ_{ij} in the sum are i.i.d. random variables with mean $\tau(\lambda)$. Then, the probability that the contact rate λ_{sd}, between the source and the destination of a randomly selected message, lies in the interval $[\lambda, \lambda + d\lambda)$, is given by

$$P\{\lambda \leq \lambda_{sd} < \lambda + d\lambda\} = \frac{T(\lambda)}{\sum_i \sum_j \tau_{ij}} = \frac{\sum_{\{i,j\} \in \mathcal{N}(\lambda)} \tau_{ij}}{\sum_i \sum_j \tau_{ij}} \tag{4.3}$$

We can express Eq. (4.3) as following:

$$P\{\lambda \leq \lambda_{sd} < \lambda + d\lambda\} = \frac{T(\lambda)}{\|\mathcal{N}(\lambda)\|} \cdot \frac{\|\mathcal{N}(\lambda)\|}{N(N-1)/2} \cdot \frac{N(N-1)/2}{\sum_i \sum_i \tau_{ij}}$$

where $\|\cdot\|$ denotes the cardinality of a set and $\frac{N(N-1)}{2}$ is the total number of node pairs in a network with N nodes. Let us further denote:

$$X_1 = \frac{T(\lambda)}{\|\mathcal{N}(\lambda)\|} \quad , \quad X_2 = \frac{\|\mathcal{N}(\lambda)\|}{N(N-1)/2} \quad , \quad X_3 = \frac{\sum_i \sum_i \tau_{ij}}{N(N-1)/2}$$

Applying the weak law of large numbers [47], it holds that for a *large network*[1]

$$X_1 \xrightarrow{p} \tau(\lambda) \quad \text{and} \quad X_2 \xrightarrow{p} f_\lambda(\lambda) \tag{4.4}$$

where $\xrightarrow{p}$ denotes convergence in probability.

Also, X_3 corresponds to the sample average of τ_{ij} over all disjoints sets $\mathcal{N}(\lambda)$. Thus, applying Cramér's theorem (*Theorem 6.5* in [47])[2] and using the convergence expressions of Eq. (4.4), we can get

$$X_3 \xrightarrow{p} \int_0^\infty \tau(y) f_\lambda(y) dy = E[\tau(\lambda)] = C$$

Similarly, using Cramér's theorem, it can be shown that the expression $X_1 \cdot X_2 \cdot \frac{1}{X_3}$ converges too, i.e.

$$X_1 \cdot X_2 \cdot \frac{1}{X_3} \xrightarrow{p} \tau(\lambda) \cdot f_\lambda(\lambda) \cdot \frac{1}{C}$$

Finally, denoting the probability density function of the source-destination contact rate λ_{sd} as $f_\tau(\lambda)$, i.e. $P\{\lambda \leq \lambda_{sd} < \lambda + d\lambda\} = f_\tau(\lambda) d\lambda$ gives us the desired result. □

As Proposition 1 shows, the source-destination contact rate distribution depends both on the contact rate distribution $f_\lambda(\lambda)$ and the traffic patterns $\tau(\lambda)$ (i.e. *joint effect of mobility and traffic*). Specifically, the probability that the contact rate of a selected node pair takes a certain value, e.g. $\lambda_{sd} \in [x, x + dx)$, is proportional to the number of pairs that contact with rate $\lambda_{ij} \in [x, x + dx)$ (i.e. $\propto f_\lambda(x)$) and the average traffic demand between them (i.e. $\propto \tau(x)$).

4.3.1 End-to-end Delivery Performance

An opportunistic routing protocol tries to deliver the end-to-end traffic demand τ_{ij}, and we would like to consider the effects of different contact patterns f_λ and traffic patterns $\tau(\lambda)$ on its performance. There exists a very large abundance of proposed schemes [131] and it would not be possible, nor would it provide any intuition, to analyze the effect of heterogeneity on each and every one. Instead, we focus here on some basic mechanisms to gain intuition.

[1] When $N \to \infty$, then $d\lambda = O\left(\frac{1}{N}\right) \to 0$, and $\|\mathcal{N}(\lambda)\| = O\left(\frac{N(N-1)}{2} d\lambda\right) = O(N) \to \infty$.

[2] Equivalently, one could use here the *Continuous Mapping Theorem*.

The approach with the minimum overhead and complexity is *Direct Transmission ("DT")*: nodes wishing to exchange data or information with each other, may do so, only when they are in direct contact, without involving any relays. For instance, DT is often assumed in content-centric applications, where a node interested in some content will query directly encountered nodes for content of interest, and retrieve it only if it is available there. Furthermore, it is the only feasible approach if nodes do not have incentives to relay traffic they are not personally interested in, e.g. due to privacy or resource-related concerns [74]. Nevertheless, DT is known to suffer from long delays and low throughput [44].

To improve the performance of direct transmission, *replication* or *relay-assisted* schemes can be used. Extra copies can be handed over to encountered nodes, and the destination can receive the message from either the source or any of the relays, reducing thus the expected delivery delay. Taken to the extreme, schemes like epidemic routing [137] forward the message at every possible encounter (deterministically, probabilistically, or based on some utility-function). Yet these do not usually scale well beyond networks with few tens of nodes, due to large resource usage. Instead, few relays are normally used, in an attempt to strike a good tradeoff.

In networks with homogeneous mobility and traffic, it is known that using just a few extra copies leads to significant performance gains. For example, in a network of 1000 nodes, simply distributing 10 extra copies to the first 10 nodes encountered provides an almost 10-fold improvement in delay compared to direct transmission [129]. Although this also comes with a 10-fold increase in the amount of (storage and bandwidth) resources needed, it presents a very useful tradeoff to DTN protocol designers.

However, when it comes to heterogeneous mobility and traffic, Proposition 1 suggests that, unlike the above example, the source is no longer equivalent with other random relays, in terms of their probability of contacting an intended destination soon. It is thus of particular interest to examine whether the above trade-off still holds, if one considers the joint effect of realistic mobility and communication traffic patterns.

We thus consider, in the following, *Relay-assisted* routing, which is a simple abstraction of schemes that use extra *randomly* chosen relays[3]. To compare the performance of Relay-assisted routing and Direct Transmission, in terms of delivery delay and delivery probability (the two main metrics considered in related work), we first define the following metrics:

(a) Delay Ratio, R: the ratio of the expected delivery delay of Relay-Assisted routing, $E[T_R]$, over the expected delivery delay of Direct Transmission routing, $E[T_{DT}]$, i.e.

$$R = \frac{E[T_R]}{E[T_{DT}]}$$

(b) Source Delivery Probability, $\mathbf{P_{(src.)}}$: the probability that a message is delivered to the destination by the source node, rather than by any of the relays.

Both metrics contain information about the performance gain of Relay-assisted routing compared to Direct Transmission. Specifically, R shows how faster (on average) a message can be delivered under Relay-assisted routing, whereas $P_{(src.)}$ gives the probability that any of the relays will actually contribute in the delivery process. It is easy to see that (i) R and $P_{(src.)}$ always take values in the interval $[0, 1]$, and (ii) the higher their values are, the less the gain due to relay nodes is.

For instance, when $R = 0.1$ Relay-assisted routing delivers (on average) a message 10 times faster than Direct Transmission, while a value $R = 0.5$ denotes that Relay-assisted routing is

[3]We will briefly consider *mobility-aware* schemes in Section 4.5.

only 2 times faster. Respectively, when $P_{(src.)} = 0.1$ the probability that the source node s meets the destination d, before any other relay node meets d, is 10%, and $P_{(src.)} = 0.5$ means that this probability is 50%. In the limiting cases, when $R, P_{(src.)} \to 1$ the message is delivered to the destination by the source node itself, while when $R, P_{(src.)} \to 0$ delivery takes place (entirely) due to the relays.

In Result 9, we derive analytical expressions for these two metrics, R and $P_{(src.)}$.

Result 9. *When Relay-assisted routing with L extra copies is considered, then*

$$R = \frac{1}{E\left[\frac{\tau(\lambda)}{\lambda}\right]} \cdot \int_0^\infty \int_0^\infty \frac{\tau(x)}{x+y} \cdot f_\lambda(x)dx \cdot f_R(y)dy$$

$$P_{(src.)} = \frac{1}{E[\tau(\lambda)]} \cdot \int_0^\infty \int_0^\infty \frac{x \cdot \tau(x)}{x+y} \cdot f_\lambda(x)dx \cdot f_R(y)dy$$

where the expectations are taken over f_λ and $f_R = f_\lambda^{(*L)}$ is the L-fold convolution of f_λ.

Proof.
Delay Ratio, **R**

Let $I_{sd}(t)$ be an indicator random variable that is equal to 1 if nodes s and d are within transmission range at time t, and 0 otherwise. Let further T_{sd} denote the random inter-contact time between node pair $\{s, d\}$:

$$T_{sd} = \inf\{t > 0 : I_{sd}(0) = 1, I_{sd}(0^+) = 0, I_{sd}(t) = 1\}.$$

Since we have assumed that contact duration is negligible for the networks we consider (Def.2), the contact process is essentially a point process, and the above could be simplified to $T_{sd} = \inf\{t > 0 : I_{sd}(0) = 1, I_{sd}(t) = 1\}$.

Assume now that end-to-end messages between $\{s, d\}$ are generated at random times and *independently* from the contact process. If T_{DT} denotes the delay of directly transmitting a message from s to d, and the contact rate between s and d is $\lambda_{sd} = x$, then one can use renewal-reward theory [121] to show that

$$E[T_{DT}|\lambda_{sd} = x] = E[T_{sd}^{(e)}|\lambda_{sd} = x] = \frac{1}{x}.$$

That is, the expected delay of direct transmission is equal to the mean of the *residual* (or *excess*) inter-contact time $T_{sd}^{(e)}$, which is an exponential variable with the same rate x.

Using the property of conditional expectation and the distribution of λ_{sd} (Proposition 1) we can get:

$$E[T_{DT}] = \int_0^\infty E[T_{DT}|\lambda_{sd} = x]f_\tau(x)dx = \int_0^\infty \frac{1}{x} f_\tau(x)dx = \frac{1}{C}\int_0^\infty \frac{\tau(x)}{x} f_\lambda(x)dx$$
$$= \frac{1}{E[\tau(\lambda)]} \cdot E\left[\frac{\tau(\lambda)}{\lambda}\right] \quad (4.5)$$

Assume now that the same messages, between $\{s, d\}$ are routed using Relay-Assisted routing, with L message copies given to L relays. Let T_R^* denote the *total* delay to deliver a message using Relay-Assisted routing, T_R the remaining delay after all L copies have been distributed, T_{fwd}

the time to distribute the L copies to the L relays, and $p_{fwd} = P(T_R^* < T_{fwd})$ the probability that the message is delivered to the destination before L relay nodes have been found.

Since relays are selected randomly (e.g. [129]), $p_{fwd} = \frac{L}{N} \to 0$ for $L \ll N$. Similarly, if $L^2 \ll N$, $\frac{T_{fwd}}{T_R} \to 0$ [129]. We can thus focus only on T_R, the time after L relays have received a copy.

Denote now with $\mathcal{L}$ the set of selected relays. Using a similar argument as in the direct transmission case, we can easily show that,

$$T_R \equiv \min_{i \in \mathcal{L} \cup \{s\}} T_{id} \sim exp(X_r) \qquad \text{and} \qquad X_r = \lambda_{sd} + \textstyle\sum_{i \in \mathcal{L}} \lambda_{id} = \lambda_{sd} + X_R$$

where $X_R = \sum_{i \in \mathcal{L}} \lambda_{id}$, and the expected value of T_R will be

$$E[T_R] = \frac{1}{X_r} = \frac{1}{\lambda_{sd} + X_R}, \tag{4.6}$$

where $\lambda_{sd} \sim f_\tau$ (Proposition 1) and $X_R \sim f_R = f_\lambda^{(*L)}$, the L-fold convolution of f_λ.

Then, from Eq. (4.6) and using the property of conditional expectation, we find:

$$\begin{aligned} E[T_R] &= \int_0^\infty \int_0^\infty E\left[T_R | \lambda_{sd} = x, X_R = y\right] f_\tau(x)dx \cdot f_R(y)dy \\ &= \int_0^\infty \int_0^\infty \frac{1}{x+y} \cdot f_\tau(x)dx \cdot f_R(y)dy = \frac{1}{E[\tau(\lambda)]} \int_0^\infty \int_0^\infty \frac{\tau(x)}{x+y} \cdot f_\lambda(x)dx \cdot f_R(y)dy \end{aligned} \tag{4.7}$$

where in the last equality we substituted the expression for f_τ from Proposition 1.

Finally, dividing Eq. (4.7) with Eq. (4.5) gives the expression of Result 9 for the delay ratio R.

Source Delivery Probability, $\mathbf{P_{(src.)}}$

Using similar arguments and notation as above, the event of the *message delivery by the source* is equivalent to the destination contacting the source before any other relay.

Then, $P_{(src.)} \equiv P\{T_{sd} < T_{r-d}\}$ (where $T_{r-d} = \min_{i \in \mathcal{L}}\{T_{id}\}$), will be given by the ratio $\frac{\lambda_{sd}}{\lambda_{sd}+X_R}$ [121]. Conditioning on the the rates λ_{sd} and X_R, we can write

$$\begin{aligned} P_{(src.)} \equiv P\{T_{sd} < T_{r-d}\} &= \int_0^\infty \int_0^\infty P\{T_{sd} < T_{r-d} | \lambda_{sd} = x, X_R = y\} f_\tau(x)dx f_R(y)dy \\ &= \int_0^\infty \int_0^\infty \frac{x}{x+y} \cdot f_\tau(x)dx \cdot f_R(y)dy \end{aligned}$$

and substituting $f_\tau(x)$ from Proposition 1 gives

$$P_{(src.)} = \int_0^\infty \int_0^\infty \frac{x}{x+y} \cdot \frac{\tau(x)}{E[\tau(\lambda)]} \cdot f_\lambda(x)dx \cdot f_R(y)dy$$

which is equal to the expression for $P_{(src.)}$ in Result 9. □

In addition to the main metrics considered in this paper (Result 9), and for ease of reference, in Table 4.1 we provide expressions for the absolute performance (message delivery delay and delivery probability) of Direct Transmission and Relay-Assisted routing. The expressions follow straight from the proof of Result 9 or through similar analysis.

Table 4.1: Expected delivery delay and delivery probability of Direct Transmission and Relay-Assisted routing.

Direct Transmission	**Relay-Assisted**
Generic Case:	
$E[T_{DT}] = \frac{1}{E[\tau(\lambda)]} \cdot E\left[\frac{\tau(\lambda)}{\lambda}\right]$	$E[T_R] = \frac{1}{E[\tau(\lambda)]} \cdot \int_0^\infty \int_0^\infty \frac{\tau(x)}{x+y} \cdot f_\lambda(x)dx \cdot f_R(y)dy$
$P\{T_{DT} \leq t\} = 1 - \frac{E[\tau(\lambda) \cdot e^{-\lambda \cdot t}]}{E[\tau(\lambda)]}$	$P\{T_R \leq t\} = 1 - \frac{E[\tau(\lambda) \cdot e^{-\lambda \cdot t}]}{E[\tau(\lambda)]} \cdot \int_0^\infty e^{-y \cdot t} \cdot f_R(y)dy$
Mobility $f_\lambda(x) \sim \Gamma(x; \alpha, \beta)$, Traffic $\tau(x) = c \cdot x^k$:	
$E[T_{DT}] = \frac{1}{\mu_\lambda} \cdot \frac{1}{1 + (k-1) \cdot CV_\lambda^2}$	$E[T_R] \geq \frac{1}{\mu_\lambda} \cdot \frac{1}{1 + k \cdot CV_\lambda^2 + L}$
$P\{T_{DT} \leq t\} = 1 - \left(1 + \mu_\lambda \cdot CV_\lambda^2 \cdot t\right)^{-\frac{1+k \cdot CV_\lambda^2}{CV_\lambda^2}}$	$P\{T_R \leq t\} = 1 - \left(1 + \mu_\lambda \cdot CV_\lambda^2 \cdot t\right)^{-\frac{1+k \cdot CV_\lambda^2 + L}{CV_\lambda^2}}$

4.3.2 Insights for Real Mobile Social Networks

The expressions we derived in Result 9 are generic and can be used under any mobility and traffic pattern (i.e. for any f_λ and $\tau(\cdot)$). However, they do not give a good feel as to how exactly these metrics are affected by mobility and traffic heterogeneity. To obtain some further insights, in this section, we consider specific classes of mobility and traffic patterns that capture commonly observed characteristics of real networks. For these classes, we derive *simple closed form expressions* that bound the performance metrics R and $P_{(src.)}$.

Mobility

We will assume the contact rates to be *gamma distributed*, i.e. $f_\lambda(x) \sim \Gamma(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$.

Our choice is initially motivated by the findings of Passarella et al. [109], who have shown, through statistical analysis of pervasive social networks' datasets, that the *Gamma distribution* matches well the observed contact rates. In addition, the analytical findings of [109], further suggest that the choice of a Gamma distribution can be supported in real MSNs and can explain many of the observed properties (e.g. distribution of *aggregate* inter-contact times). Finally, by selecting appropriately the parameters α and β of a Gamma distribution, we can assign *any* desired value to the mean value μ_λ and the variance σ_λ^2 of the contact rates[4]. This allows us to describe (or fit up to the first two moments) a large range of scenarios with different mobility heterogeneities captured by $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$.

Traffic

We further describe the traffic using a *polynomial function* of the form $\tau(x) = c \cdot x^k$, $c > 0$.

As in the case of mobility, the reasons for our choice are as following. Observations of real networks have shown that the nodes with high contact frequencies tend to exchange more traffic [54, 138], which is consistent with the above choice when $k > 0$. Second, the exact traffic patterns (i.e. $\tau(x)$) in a real scenario are difficult (if not impossible) to determine, and, hence,

[4]The mean value and variance of a gamma distribution are given by $\mu_\lambda = \frac{\alpha}{\beta}$ and $\sigma_\lambda^2 = \frac{\alpha}{\beta^2}$, respectively.

it is more probable that simple methods will be used. For example, one might get some traffic samples and perform linear regression on the measured data. This would result in a linear $\tau(x)$ (i.e. $k = 1$). Our model extends this logic by going beyond linear fitting and allowing as well sub- and super-linear fitted traffic patterns. In general, the values of k capture the amount of traffic heterogeneity. Furthermore, by choosing $0 < k < 1$ (or $k > 1$) one can emulate concave (or convex) functions and, thus, approximate different traffic patterns. Finally, one can also consider negative correlations, by choosing $k < 0$. Although less common, these could arise, for example, in applications where users want to communicate more when they do not meet frequently (e.g. messaging).

Under the above assumptions, the following result for the relative performance of the information dissemination mechanisms we consider in this paper, holds. The corresponding expressions for the absolute performance metrics are given in Table 4.1.

Result 10. *In a Heterogeneous Contact Network where $f_\lambda \sim \Gamma(\alpha, \beta)$ with mean value μ_λ and variance σ_λ^2 (coefficient of variation $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$) and $\tau(x) = c \cdot x^k$, it holds:*

$$1 \geq \quad R \quad \geq \quad R_{min} = \frac{1 + (k-1) \cdot CV_\lambda^2}{1 + k \cdot CV_\lambda^2 + L} \tag{4.8}$$

for $k > k_{min} = 1 - \frac{1}{CV_\lambda^2}$, and

$$1 \geq \quad P_{(src.)} \quad \geq \quad P_{min} = \frac{1 + k \cdot CV_\lambda^2}{1 + (k+1) \cdot CV_\lambda^2 + L} \tag{4.9}$$

for $k > k_{min} = -\frac{1}{CV_\lambda^2}$.

Proof.
Delay Ratio, **R**

The expression for the delay ratio R of Result 9 can be written as

$$R = \frac{1}{E\left[\frac{\tau(\lambda)}{\lambda}\right]} \cdot \int_0^\infty E_R\left[\frac{1}{x+y}\right] \cdot \tau(x) \cdot f_\lambda(x) dx \tag{4.10}$$

where the expectation $E_R[\cdot]$ is taken over f_R. Using Jensen's inequality[5] for the function $h(y) = \frac{1}{x+y}$, we get:

$$E_R\left[\frac{1}{x+y}\right] \geq \frac{1}{x + E_R[y]} \tag{4.11}$$

where $E_R[y]$ is given by (as the expectation of a sum of L i.i.d. random variables with expectation μ_λ) [121]:

$$E_R[y] = E[X_R] = E\left[\sum_{i \in \mathcal{L}} \lambda_{id}\right] = L \cdot \mu_\lambda \tag{4.12}$$

Hence, using Eq. (4.11) and Eq. (4.12) in Eq. (4.10), we get

$$R \geq \frac{1}{E\left[\frac{\tau(\lambda)}{\lambda}\right]} \cdot \int_0^\infty \frac{\tau(x)}{x + L \cdot \mu_\lambda} \cdot f_\lambda(x) dx = \frac{1}{E\left[\frac{\tau(\lambda)}{\lambda}\right]} \cdot E\left[\frac{\tau(\lambda)}{\lambda + L \cdot \mu_\lambda}\right] \tag{4.13}$$

[5] Jensen's inequality for a convex function $h(x)$: $E[h(x)] \geq h\left(E[x]\right)$.

Now, Eq. (4.13), for $\tau(x) = c \cdot x^k$, is written as

$$R \geq \frac{1}{E[\lambda^{k-1}]} \cdot E\left[\frac{\lambda^k}{\lambda + L \cdot \mu_\lambda}\right] \tag{4.14}$$

The expectations in Eq. (4.14) are taken over the contact rates'(*Gamma*) distribution, whose general form is [121]

$$f_\lambda(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

where $\alpha > 0$ is the *shape* parameter, $\beta > 0$ the *rate* parameter. Its mean value and variance are given by $\mu_\lambda = \frac{\alpha}{\beta}$ and $\sigma_\lambda^2 = \frac{\alpha}{\beta^2}$, respectively, and, equivalently, we can write

$$\alpha = 1/CV_\lambda^2, \qquad \beta = 1/\left(\mu_\lambda \cdot CV_\lambda^2\right) \tag{4.15}$$

To calculate Eq. (4.14), first we find an expression for $E[\lambda^{k-1}]$:

$$\begin{aligned} E[\lambda^{k-1}] &= \int_0^\infty x^{k-1} f_\lambda(x) dx = \int_0^\infty x^{k-1} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\ &= \frac{\Gamma(k-1+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^{k-1}} \int_0^\infty \frac{\beta^{k-1+\alpha}}{\Gamma(k-1+\alpha)} x^{(k-1+\alpha)-1} e^{-\beta x} dx = \frac{\Gamma(k-1+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^{k-1}} \end{aligned} \tag{4.16}$$

where the integral in the second line is equal to 1 because the integrated function is the pdf of a Gamma distribution with parameters $\alpha^{'} = k - 1 + \alpha$ (it must hold that $\alpha^{'} > 0$, which means that $k > 1 - \alpha = 1 - \frac{1}{CV_\lambda^2}$) and $\beta^{'} = \beta$.

Similarly to the derivation of Eq. (4.16), it can be shown that

$$\begin{aligned} E\left[\frac{\lambda^k}{\lambda + L \cdot \mu_\lambda}\right] == \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^k} \cdot \int_0^\infty \frac{1}{x + L \cdot \mu_\lambda} \frac{\beta^{k+\alpha}}{\Gamma(k+\alpha)} x^{k+\alpha-1} e^{-\beta x} dx \\ = \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^k} \cdot E_{\lambda'}\left[\frac{1}{\lambda' + L \cdot \mu_\lambda}\right] \end{aligned} \tag{4.17}$$

where λ' follows a Gamma distribution with parameters $\alpha' = k + \alpha$ and $\beta' = \beta$. Since the function $g(x) = \frac{1}{x+c}$ is convex, we can apply Jensen's inequality to Eq. (4.17) and get

$$E\left[\frac{\lambda^k}{\lambda + L \cdot \mu_\lambda}\right] \geq \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^k} \cdot \frac{1}{E[\lambda'] + L \cdot \mu_\lambda} = \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)} \frac{1}{\beta^k} \cdot \frac{1}{\frac{k+\alpha}{\beta} + L \cdot \mu_\lambda} \tag{4.18}$$

where we substituted $E[\lambda'] = \frac{\alpha'}{\beta'} = \frac{k+\alpha}{\beta}$.

Thus, from Eq. (4.16) and Eq. (4.18), it holds for R (Eq. (4.14)):

$$R \geq \frac{\Gamma(k+\alpha)}{\Gamma(k-1+\alpha)} \cdot \frac{1}{\beta} \cdot \frac{1}{\frac{k+\alpha}{\beta} + L \cdot \mu_\lambda}$$

and because of the Gamma function's property $\Gamma(z+1) = z \cdot \Gamma(z)$, we can write

$$R \geq \frac{k-1+\alpha}{\beta} \cdot \frac{1}{\frac{k+\alpha}{\beta} + L \cdot \mu_\lambda} \tag{4.19}$$

and Eq. (4.8) follows easily by substituting α and β from Eq. (4.15) to Eq. (4.19).

Source Delivery Probability, $\mathbf{P_{(src.)}}$

Using the same notation, the expression for the delivery probability $P_{(src.)}$ of Result 9 can be written as

$$P_{(src.)} = \frac{1}{E[\tau(\lambda)]} \cdot \int_0^\infty E_R\left[\frac{x \cdot \tau(x)}{x+y}\right] \cdot f_\lambda(x)dx \tag{4.20}$$

and applying Jensen's inequality as in Eq. (4.11), we get

$$P_{(src.)} \geq \int_0^\infty \frac{x \cdot \tau(x)}{x + L \cdot \mu_\lambda} \cdot f_\lambda(x)dx = \frac{1}{E[\tau(\lambda)]} \cdot E\left[\frac{\lambda \cdot \tau(\lambda)}{\lambda + L \cdot \mu_\lambda}\right] \tag{4.21}$$

Now, setting $\tau(x) = c \cdot x^k$ in Eq. (4.21), gives

$$P_{(src.)} \geq \frac{1}{E[\lambda^k]} \cdot E\left[\frac{\lambda^{k+1}}{\lambda + L \cdot \mu_\lambda}\right] \tag{4.22}$$

Using Eq. (4.16) - Eq. (4.18) (where instead of k we consider $k+1$), the result for $P_{(src.)}$ follows similarly as before. □

The expressions of Result 10 depend only on 3 parameters (CV_λ, k, L) and, thus, could be used to tune Relay-Assisted schemes: At first, since mobility (CV_λ) and traffic (k) parameters are characteristics of the network, they either remain constant or change slowly over a long time period. Hence, we can assume that nodes know their values, or can estimate them (e.g. with a distributed mechanism, locally, etc.) [8, 45]. Then, the required number of relays L to achieve a certain expected delay, could be easily estimated.
Practical Example: If the measured network characteristics are $CV_\lambda = 2$ and $k = 2$, then from Result 10 we get $R = \frac{5}{9+L}$. Therefore, to achieve delivery delay two times faster than Direct Transmission, one extra copy should be used ($L = 1 \rightarrow R = 0.5$), while to achieve 4 times faster delivery, $L = 11$ relay nodes are needed. In the latter case, if traffic/mobility heterogeneity has not been taken into account [129], the prediction would be $L = 3$ and this would lead only to 2.5 (instead of 4) times faster delivery (i.e. $R = \frac{5}{12}$).

4.3.3 Implications

It is evident from the above example that traffic heterogeneity can have a major impact on performance and thus protocol design. Table 4.2 formalizes this impact, by considering how R_{min} and P_{min} (Eq. (4.8) and Eq. (4.9)) behave:
The *middle column* shows their monotonicity as mobility heterogeneity (CV_λ), traffic heterogeneity (k), and amount of extra copies (L) increase. For instance, when k increases ($\nearrow$), R_{min} and P_{min} increase ($\nearrow$) too.
The *right column* gives their values in the limit for large/small k or CV_λ; e.g.

$$\lim_{CV_\lambda \to 0} R_{min} = \lim_{CV_\lambda \to 0} \frac{1 + (k-1) \cdot CV_\lambda^2}{1 + k \cdot CV_\lambda^2 + L} = \frac{1}{1+L}$$

and

$$\lim_{CV_\lambda \to \infty} P_{min} = \lim_{CV_\lambda \to \infty} \frac{1 + k \cdot CV_\lambda^2}{1 + (k+1) \cdot CV_\lambda^2 + L} = 1 - \frac{1}{k+1}$$

In this section, we elaborate on some important implications that follow from Table 4.2.

Table 4.2: R_{min}, P_{min}: Monotonicity and Asymptotic Limits

Parameter x	Monotonicity as parameter x increases ↗	Limits for $x \to min\{x\}$	Limits for $x \to max\{x\}$
mobility heterogeneity: $CV_\lambda \in [0, \infty)$	R_{min} increases ↗ , if $k > 1 + \frac{1}{L}$; decreases ↘ , otherwise	$\lim_{CV_\lambda \to 0} R_{min} = \mathbf{\frac{1}{1+L}}$	$\lim_{CV_\lambda \to \infty} R_{min} = \mathbf{1\text{-}\frac{1}{k}}$
	P_{min} increases ↗ , if $k > \frac{1}{L}$; decreases ↘ , otherwise	$\lim_{CV_\lambda \to 0} P_{min} = \mathbf{\frac{1}{1+L}}$	$\lim_{CV_\lambda \to \infty} P_{min} = \mathbf{1\text{-}\frac{1}{k+1}}$
traffic heterogeneity: $k \in (k_{min}, \infty)$	R_{min}, P_{min} increase ↗	$\lim_{k \to k_{min}} R_{min}, P_{min} = \mathbf{0}$	$\lim_{k \to \infty} R_{min}, P_{min} = \mathbf{1}$
extra copies: $L \quad (L \ll N)$	R_{min}, P_{min} decrease ↘	-	

Gain of Extra Copies

A strong positive correlation (large k) between traffic and mobility reduces the added value of extra copies (i.e. R_{min}, P_{min} ↗ as k ↗). This indicates that, as correlation (k) increases, one needs to distribute message copies to more relays nodes in order to achieve a certain performance improvement compared to the baseline, Direct Transmission.

In contrast, a negative (or weak positive) correlation renders each extra copy more useful (i.e. R_{min}, $P_{min} \to 0$ as $k \to k_{min}$[6]). The fact that a *weak positive* correlation, e.g. $k \in (0, \frac{1}{L})$, actually makes extra copies more useful might be a bit surprising. However, it is explained as following: Mobility heterogeneity (when traffic is homogeneous or uncorrelated with mobility) affects negatively the message delivery delay (of random protocols and Direct Transmission) [76, 113], whereas positively-correlated traffic has an opposite effect (i.e. decreases delay). The counterbalancing effects of these two factors determine a *threshold* (e.g. $1 + \frac{1}{L}$ for R_{min} or $\frac{1}{L}$ for P_{min}) under which the negative effects of heterogeneity affect more the message delivery process. Our framework, not only reveals this inherent trade-off, but also provides the tools for quantifying such thresholds.

From the above discussion it becomes evident that it is crucial to identify whether a traffic-mobility correlation exists in a given scenario, and what its nature is, as this could decide whether the overhead of using few or more extra copies is justified or would just waste a lot of valuable resources. In practice, this means that a relay-assisted protocol should be complemented with an online estimation algorithm, collaborative or local. Such schemes have been proposed [8, 45] to collect contact related information for forwarding algorithms, but would now need to maintain also traffic-related information *and* correlate it with the information about the node contact rates, in an efficient manner.

[6]The values of k_{min} are given in Result 10.

Routing for Unicast Applications

For high heterogeneity (traffic and mobility), our results imply that a unicast message is likely to arrive to its destination at the time the source and destination come *in contact* (i.e. R_{min}, $P_{min} \to 1$ as $k, CV_\lambda \to \infty$). This raises questions about the usefulness of mobile social networking for unicast applications in which end-to-end traffic is expected to be highly correlated with contact frequency (e.g. Facebook messaging) [54, 138].

On the other hand, our results suggest that potential unicast applications with an end-to-end traffic demand between nodes with non-frequent meetings, i.e. scenarios with small or negative k, (e.g. social peers residing in different communities) could benefit a lot (more than normally assumed).

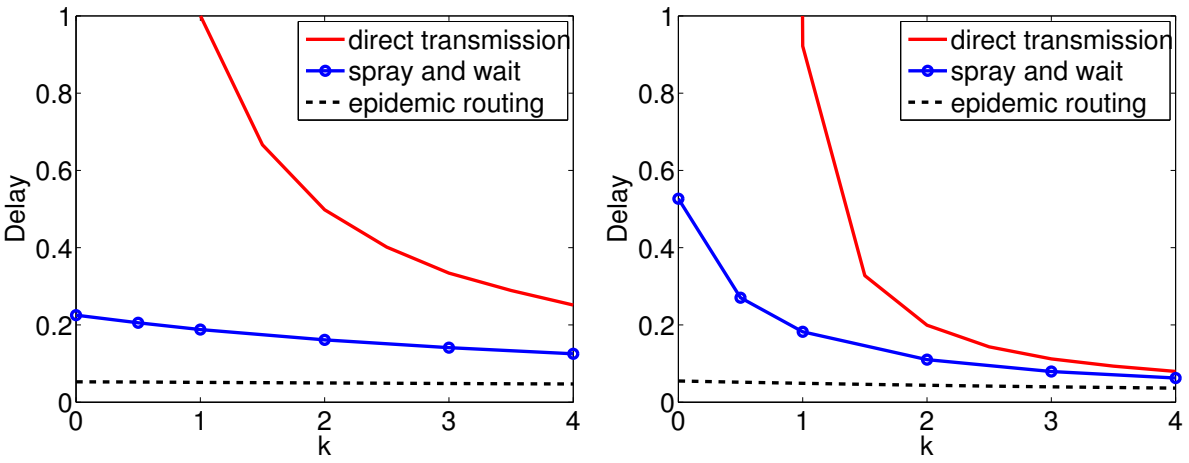
Although these observations might appear somewhat self-evident at first glance (note however the case described in the previous subsection), the question of how to tune protocols and choose the right number of replicas stills remains. To our best knowledge, our results are the first to provide closed form, *quantitative* insights into the tradeoffs involved in real scenarios with both mobility and traffic heterogeneity.

Moreover, one could raise a point about their applicability for sophisticated protocols that choose relays intelligently (e.g. based on contact rates, social graphs). In this case, a source node could try to wait and select better relays than giving the copies to the first randomly encountered peers, thus improving the impact per replica. Nevertheless, in a highly heterogeneous scenario, a source might need to wait a long time until it encounters such good relays ("spray" phase) and this could counter-balance the effect of better relays. In Section 4.5, we prove that the qualitative implications of our results hold also for such mobility-aware protocols, which exploit mobility heterogeneity in order to select better relays. A complementary explanation for this qualitative result is given in the end of this section (see Fig. 4.2 and the corresponding commentary).

Content-Centric Communication

While our results are somewhat pessimistic when it comes to the usefulness of MSNs for unicast applications, the opposite holds when it comes to modern, content-centric applications (e.g. file sharing, D2D-based offloading, service composition). In such applications nodes are looking, for example, for some content of interest [50] or service [119], which they can access directly from *any* encountered node that offers it. If the interests of nodes are heterogeneous (which is known to be the case [15]) and nodes with similar mobility patterns tend to have *some* similarity in their interests too (evidence for this does exist [135]), then our results suggest: (i) that there is a better chance to find a content or service "soon" from a directly encountered node than one would expect in homogeneous scenarios, and (ii) coming up with complex, resource-costly mechanisms, e.g. multi-hop query-response, directories, etc., might not be necessary.

To put some extra evidence on our arguments and further demonstrate *how* and *why* traffic heterogeneity affects the relative performance, in Fig. 4.2 we compare the message delay of (i) Direct Transmission (i.e. the protocol with the *highest* delay), (ii) Relay-assisted routing (Spray and Wait, *SnW*, [129] with $L = 5$ copies) and (iii) Epidemic routing [137] (i.e. the protocol with the *lowest* delay), in two scenarios, for varying traffic heterogeneity (k). Two main observations, with respect to the previous implications, can be made in Fig. 4.2.



(a) $CV_\lambda = 1$ (b) $CV_\lambda = 2$

Figure 4.2: Message delay under Direct Transmission, Spray and Wait ($L = 5$), and Epidemic routing in scenarios with varying traffic heterogeneity; mobility parameters are $\mu_\lambda = 1$ and (a) $CV_\lambda = 1$ and (b) $CV_\lambda = 2$.

At first, an increasing amount of traffic heterogeneity/correlation closes the performance gap between the best (Epidemic) and the worst (Direct Transmission) forwarding policies. Hence, it becomes evident that the possible gain one could achieve by using *any routing protocol* and *any number of extra copies*, diminishes. As a result, routing schemes, whose design is crucial in homogeneous scenarios (since the improvement gap is large; see Fig. 4.2 for regions with low k), become less important in heterogeneous scenarios with highly correlated traffic (since the improvement cannot be large; see Fig. 4.2 for regions with high k) and/or less necessary (since comparable performance can be achieved with Direct Transmission; e.g. Fig. 4.2(b) for $k = 4$).

Second, the delay of Direct Transmission decreases radically as traffic heterogeneity increases[7]. Although the delay of Relay-assisted routing decreases with traffic heterogeneity k too, the effect is less significant. Specifically, an observation of the delay curves for Direct Transmission and Relay-assisted routing in Fig. 4.2(a), shows that the delay ratio $R = \frac{E[T_R]}{E[T_{DT}]}$ increases as traffic becomes more heterogeneous. However, this increase is mainly due to the *improved performance of Direct Transmission* rather than this of Relay-assisted routing.

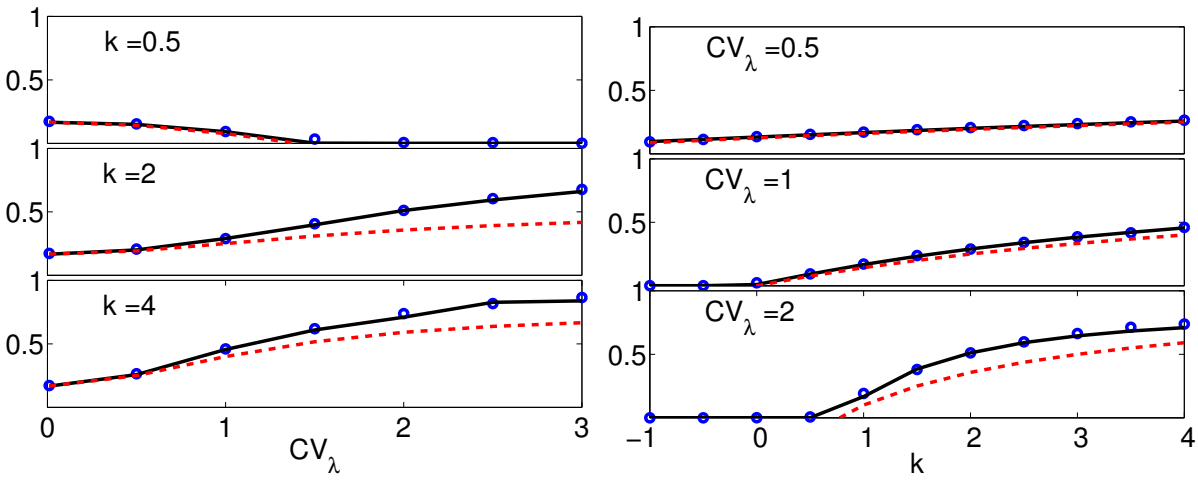
4.4 Model Validation

To validate our model and analysis, in this section we compare the theoretical results against Monte Carlo simulations on various synthetic scenarios, and on datasets of real networks.

4.4.1 Synthetic Simulations

We generate synthetic networks, conforming to the mobility and traffic models of Section 4.2, as following:

[7]The convergence is faster for scenarios where node mobility is more heterogeneous (Fig. 4.2(b)), suggesting, thus, that the effects of traffic heterogeneity are even more important when coupled with highly heterogeneous node mobility.



(a) $R = \frac{E[T_R]}{E[T_{DT}]}$ vs. CV_λ (b) $R = \frac{E[T_R]}{E[T_{DT}]}$ vs. k

Figure 4.3: R in scenarios with varying (a) mobility and (b) traffic heterogeneity. Simulation results are denoted with circles; the theoretical predictions of Result 9 (exact predictions) with continuous lines; and the lower bounds R_{min} (Result 10) with dashed lines.

(i) We assign to each pair $\{i, j\}$ a contact rate λ_{ij}, which we draw randomly from f_λ, and create a sequence of contact events (Poisson process with rate λ_{ij}).
(ii) Since $E[\tau_{ij}] = \tau(\lambda_{ij})$ (from Def. 9), we draw the traffic rate for each pair $\{i, j\}$ as $\tau_{ij} \sim Uniform[0, 2 \cdot \tau(\lambda_{ij})]$.
(iii) Then, we simulate a large number of message exchanges, choosing randomly for each message the source-destination pair according to the weights τ_{ij}.

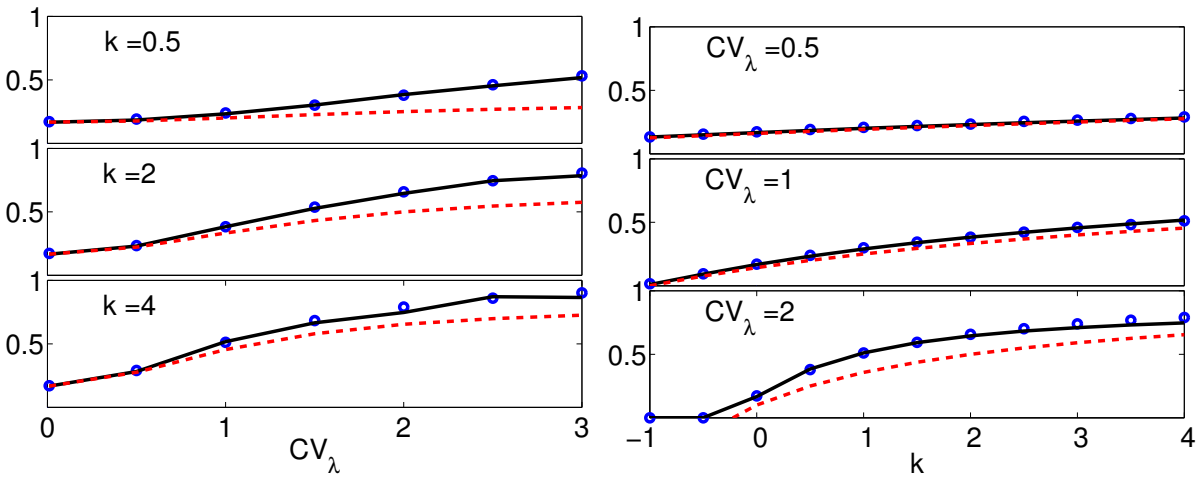
We created different scenarios $(N, L, f_\lambda, \tau(\cdot))$ to verify our analysis under various network parameters. Here, we present the simulation results for scenarios with $N = 500$ nodes[8]. As Relay-assisted routing, we used the *Spray and Wait* protocol [129] with $L = 5$ copies. To be consistent with the analysis of Section 4.3.2, we used the *Gamma distribution* as the contact rates distribution f_λ and traffic functions of polynomial form, $\tau(x) = c \cdot x^k$.

In Fig. 4.3 and Fig. 4.4 we present simulation results for the ratios R and probabilities $P_{(src.)}$, along with the corresponding theoretical results (exact predictions of Result 9 and lower bounds of Result 10), in scenarios with varying mobility and traffic heterogeneity.

Fig. 4.3 shows the *delay ratio* R: (a) in three scenarios with different traffic functions $\tau(x)$ (namely[9]: $c \cdot \sqrt{x}$, $c \cdot x^2$, and $c \cdot x^4$), under varying *mobility* heterogeneity; and (b) in three mobility scenarios with $CV_\lambda = \{0.5, 1, 2\}$, under varying *traffic* heterogeneity. A first observation is that the exact expressions of Result 9 (continuous lines) can accurately predict the metric R (simulation results are denoted with circles). Additionally, the lower bounds are always below the simulation curves (as expected), and in many scenarios are quite tight.

Under the same mobility (CV_λ) and traffic (k) simulation scenarios, similar observations can be made for the *source delivery probability* $P_{(src.)}$ in Fig. 4.4, where the exact expressions of Result 9 accurately match the simulation results and the bounds of Result 10 are tight in most scenarios.

[8]The simulations we ran for networks with $N \in [100, 1000]$ nodes, gave us similar results.
[9]The value of c does not affect the performance (see also Result 10).



(a) $P_{(src.)}$ vs. CV_λ (b) $P_{(src.)}$ vs. k

Figure 4.4: $P_{(src.)}$ in scenarios with varying (a) mobility and (b) traffic heterogeneity. Simulation results are denoted with circles; the theoretical predictions of Result 9 (exact predictions) with continuous lines; and the lower bounds P_{min} (Result 10) with dashed lines.

In general, for both the metrics R and $P_{(src.)}$, the theoretical lower bounds are less tight for scenarios where mobility is quite heterogeneous. Specifically, in Fig. 4.3(a) and 4.4(a), the bounds are less close to the simulation curves in the regimes where CV_λ becomes larger than 2. Also, in the scenarios with varying traffic heterogeneity (Fig. 4.3(b) and 4.4(b)), the bounds are tight for scenarios with small and moderate mobility heterogeneity, and become less tight only in the scenarios with $CV_\lambda = 2$ (bottom plots of Fig. 4.3(b) and 4.4(b)).

In every scenario, the simulation curves R and $P_{(src.)}$ have the monotonicity we predicted in Table 4.2 (middle column) for the theoretical bounds R_{min} and P_{min}. For instance, when traffic heterogeneity (k) increases, R and $P_{(src.)}$ always increase as well (Fig. 4.3(b) and 4.4(b)). Also, in the regimes that $k \leq k_{min}$[10] the simulation values of the considered metrics become almost zero, and for large k (especially in the bottom plots of Fig. 4.3(b) and 4.4(b), where mobility is also very heterogeneous) they get close to 1, thus validating the qualitative predictions of Table 4.2 (right column).

The simulation results in Fig. 4.3(a) and 4.4(a), where we present scenarios with varying mobility heterogeneity (CV_λ), validate our predictions for the monotonicity and limiting behavior as well. For example, in Fig. 4.3(a) for $k = 0.5$, where the traffic-mobility correlation is small (the same holds also for negative correlations), R and R_{min} decrease as the mobility heterogeneity increases (as suggested in Table 4.2). In the rest of the plots, the bounds and the corresponding simulated values increase, demonstrating that the gain of the extra copies diminishes under such conditions, and, thus, confirming our qualitative results (Section 4.3.3). For example, in the bottom plot ($k = 4$) of Fig. 4.3(a), we can see that the improvement offered by the extra relays is at most 6× (since $R = \frac{1}{1+L} = \frac{1}{6}$) for homogeneous network ($CV_\lambda = 0$), while for $CV_\lambda > 2$ the extra gain is at most 1.25× (since $R > 0.8$); that is, even using 5 relays will only

[10](i) $k_{min} = 0$ and $k_{min} = 0.75$ for the middle ($CV_\lambda = 1$) and bottom ($CV_\lambda = 2$) plots in Fig. 4.3(b), respectively; and (ii) $k_{min} = -1$ and $k_{min} = -0.25$ for the middle ($CV_\lambda = 1$) and bottom ($CV_\lambda = 2$) plots in Fig. 4.4(b), respectively.

Table 4.3: Datasets Information

Dataset	Nb of Nodes	Contacts	Traffic
Gowalla/Twitter	(AU) 1004	Check-ins	Tweets
	(SF) 479		
Strathclyde	24	Bluetooth Proximity	Calls/SMS

marginally improve the delay. Similarly, from Fig. 4.4(b) and for $CV_\lambda = 2$, we can see that, while for almost homogeneous traffic ($k < 0.5$) the probability of the message being delivered through direct transmission, $P_{(src.)}$, gets less than 40%, when traffic becomes very heterogeneous ($k \geq 4$), this probability is around 80%.

4.4.2 Real-World Networks

To further investigate the applicability of our results in real-world networks, we conduct simulations on datasets collected from online social networks (*Gowalla / Twitter* dataset [54]) and a mobile phone usage experiment (*Strathclyde* dataset [90]). In the following discussion we present the datasets, whose main features can be found also in Table 4.3.[11]

Gowalla / Twitter dataset

Gowalla was a location-based social network, where users were able to *check-in* at "spots" (bars, shops etc.) through their mobile phones. In addition, a user could connect her Gowalla account to her Twitter account. Hence, from this dataset, we could retrieve information related both to nodes' mobility (Gowalla *check-ins*) and communication traffic (*tweets*).

Mobility: In this dataset, we consider as a contact event the time when two users reside in the same "spot" simultaneously[12]. The contact rates λ_{ij} can be computed from the number of the contact events and the inter-contact time intervals. Then, to incorporate this information in our model, we fit the contact rates distribution f_λ with a known probability distribution $\hat{f}_\lambda$. Specifically, in the two cities, Austin (AU) and San Francisco (SF), for which we have the most user records (1004 and 479 nodes, respectively), the experimental CCDF (*complementary cumulative distribution function*) of the contact rates λ_{ij} can be approximated by a straight line on a log-log plot. This implies that f_λ could be fitted with a *Pareto* distribution, instead of the Gamma distribution assumed in Section 4.3.2 and often observed in traces. Therefore, we use here the expressions of Result 9, instead of Result 10.

Communication Traffic: As an indication for the communication traffic that two nodes would exchange in an opportunistic network, we use the number of *tweets* in which they are both involved. Hence, for each pair $\{i, j\}$ we set its traffic rate τ_{ij} equal to the number of tweets posted by i to j or by j to i, i.e. $\tau_{ij} = \#tweets_{ij}$. Then, we approximate the observed

[11] Here, we need to stress that the selected datasets are not necessarily characteristic examples of MSNs; e.g., Gowalla is a very sparse dataset in terms of node contacts, and phone calls (*Strathclyde*) is not considered among the main opportunistic applications. However, they are some of the few available datasets containing the type of data we needed (i.e. both mobility and traffic information), and, thus, this was our best option for a realistic validation.

[12] Since Gowalla users only check in and do not check out, we cannot infer directly this information. Therefore, following the methodology of [54], we assumed that each user remains at a spot she visited for 1 hour.

relation between traffic and contact rates ($\tau_{ij} \sim \lambda_{ij}$) with a function $\hat{\tau}(x)$, in order to use it in our theoretical expressions. We also investigate more possible correlations between the opportunistic traffic (τ_{ij}) and the Twitter traffic ($\#tweets$), by creating two additional scenarios where we set $\tau_{ij} = \sqrt{\#tweets_{ij}}$ and $\tau_{ij} = (\#tweets_{ij})^2$. The approximative functions $\hat{\tau}(x)$ for each scenario are presented in Table 4.4, where we can see $\hat{\tau}(x)$ being of type $c \cdot x^k$ with $k < 1$.

Strathclyde dataset

The Strathclyde dataset was collected in an experiment, in which 24 high school students were selected and given modified smartphones, which recorded proximity events (through Bluetooth), calls and sms exchanged between the phone user and the other participants.

Mobility: In this dataset the contact events were already recorded and, thus, we did not have to preprocess the data as in the Gowalla dataset. We followed the same methodology to calculate the contact rates λ_{ij} and fit their distribution with a *Gamma* distribution, denoted as $\hat{f}_\lambda$.

Communication Traffic: We consider three scenarios, in each of which we use a different communication traffic metric: (i) total number of calls and sms, $\tau_{ij} = \#calls_{ij} + \#sms_{ij}$, (ii) total duration of calls, $\tau_{ij} = callTime_{ij}$, and (iii) total length of sms (in characters), $\tau_{ij} = smsLength_{ij}$. For each scenario, we fit function $\hat{\tau}(x)$ as before, through the relation $\tau_{ij} \sim \lambda_{ij}$.

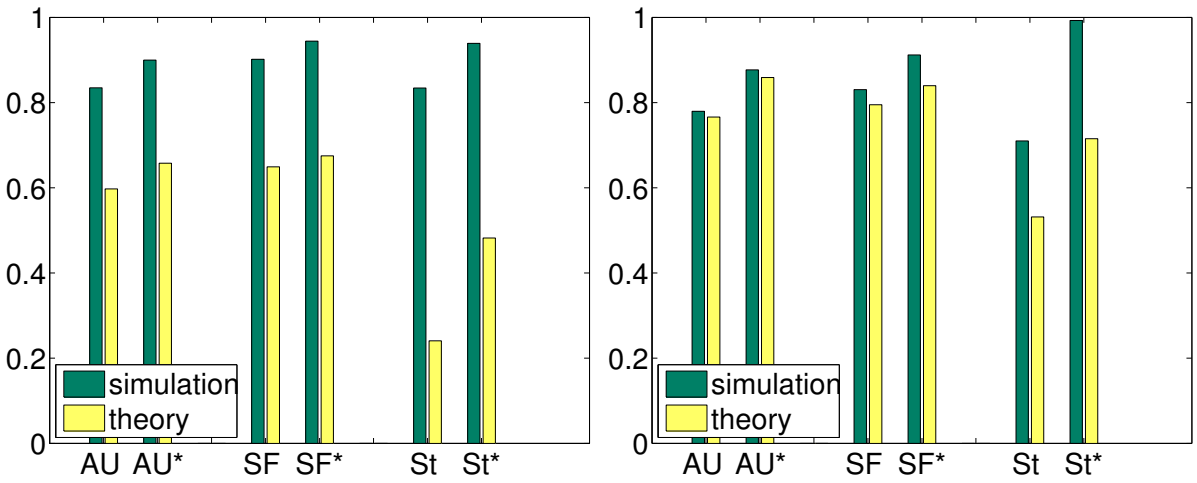
Simulations

In both datasets and for each traffic scenario, we generate 10000 messages at random time points, choosing each time the source - destination pair according to the weights τ_{ij}. We consider Direct Transmission and Spray and Wait routing [129] with $L = 2, 5, 10, 20$ copies per message. In the analytical expressions we use the fitted functions $\hat{f}_\lambda(x)$ and $\hat{\tau}(x)$.

In Fig. 4.5 we present the simulation values for the ratio R and the probability $P_{(src.)}$ (green/left bars), and the corresponding theoretical predictions (yellow/right bars). We consider *homogeneous* and *heterogenous* (denoted with $*$) traffic scenarios in the Gowalla/Twitter (AU and SF) and Strathclyde (St) datasets. The first observation is that in all scenarios, for heterogeneous traffic (i.e. scenarios denoted with $*$), the values of the metrics R and $P_{(src.)}$ increase, compared to the corresponding homogeneous scenarios. This shows that the relative gains of relay-assisted schemes decrease with traffic heterogeneity, as our theoretical results predict. Moreover, larger performance differences predicted by our theory, are matched by larger performance differences in the respective simulation scenarios as well. For example, in the SF scenarios (middle bars in Fig. 4.5(a) and Fig. 4.5(b)), the theoretical predictions for heterogeneous traffic are slightly higher than for the homogeneous case; the same holds also for the

Table 4.4: Fitting traffic functions for the Gowalla dataset

Scenarios:	S1	S2	S3
τ_{ij}	$\sqrt{\#tweets_{ij}}$	$\#tweets_{ij}$	$(\#tweets_{ij})^2$
$\hat{\tau}(x)$ (AU)	$c \cdot x^{0.6}$	$c \cdot x^{0.83}$	$c \cdot x^{0.79}$
$\hat{\tau}(x)$ (SF)	$c \cdot x^{0.31}$	$c \cdot x^{0.35}$	$c \cdot x^{0.37}$



(a) $R = \frac{E[T_R]}{E[T_{DT}]}$ (b) $P_{(src.)}$

Figure 4.5: Simulation results for R and $P_{(src.)}$ and theoretical predictions for homogeneous and heterogeneous ($*$) traffic scenarios on the datasets.

simulation results, where it can be seen that R and $P_{(src.)}$ do not significantly increase with traffic heterogeneity. On the other hand, in the *St* scenarios (right bars in Fig. 4.5(a) and Fig. 4.5(b)), our results predict a higher difference (between heterogeneous and homogeneous cases) than before, which is also confirmed by the simulation results where the performance effects are not negligible.

To further demonstrate to what extent our results can capture the effect of traffic heterogeneity in real scenarios, in Table 4.5 we focus on the qualitative predictions of our theory, by comparing a number of scenarios with different amounts of heterogeneity to each other, for the Gowalla/Twitter dataset[13]. Specifically, if the simulated performance improves from one scenario to another, and so is the theoretical prediction, the prediction is assumed to be correct and denoted with ✓. "Incorrect" predictions are denoted with ×. For example, in the scenarios *AU-S1* and *SF-S3* the simulation values for the ratios R are $R^{(AU-S1)} = 0.89$ and $R^{(SF-S3)} = 0.94$, i.e. $R^{(AU-S1)} < R^{(SF-S3)}$. For the theoretical predictions it holds also that $R^{(AU-S1)} = 0.64 < R^{(SF-S3)} = 0.68$ and, thus, the prediction is assumed to be correct. The elements above the diagonal refer to the ratio R, whereas the lower triangular part refers to the probability $P_{(src.)}$ predictions.

It is evident that in the majority of the cases we consider, the theoretical results can capture the relative changes in network performance, even between different environments (i.e. between *AU* and *SF*)[14]. The same conclusions can be reached by the analysis in the Starthclyde dataset, in which *all* the respective comparisons were found to be correct ✓.

[13]We denote with *S1*, *S2* and *S3* the corresponding scenarios presented in Table 4.4 and with *HOM* the scenarios with homogeneous traffic.

[14]Differences in simulation and theoretical results between different heterogeneous scenarios of the same traces, are very small (due to the dataset limitations), and that is also the main reason for some × entries in Table 4.5.

Table 4.5: Comparison of predictions for the metrics R and $P_{(src.)}$ between different scenarios on the Gowalla dataset

*	R	AU			SF		
$P_{(src.)}$	*	HOM	S1	S2	HOM	S1	S3
AU	HOM	*	✓	✓	✓	✓	✓
	S1	✓	*	✓	✓	✓	✓
	S2	✓	✓	*	×	✓	✓
SF	HOM	✓	✓	✓	*	✓	✓
	S1	✓	✓	×	✓	*	✓
	S3	✓	✓	×	✓	✓	*

4.5 Extensions

We have tried to present our results in the context of simple schemes (e.g. unicast traffic, random relay selection), to keep analysis tractable and illustrate key principles. In this section, we discuss how our framework could be applied in some additional cases. Although far from complete, we believe this set of examples, further underlines the utility of our analysis.

4.5.1 Mobility-Aware Protocols

Mobility-aware schemes are often used to select good relays for the intended replicas, rather than picking random ones, e.g. [29, 59, 60, 98]. The selection of the relays is usually based on their social or mobility characteristics. For instance, in *encounter-based routing (EBR)* [98], the more frequently a node i encounters node d, the higher the probability to become a relay of a message destined to d.

The relay-selection mechanism in a number of proposed mobility-aware protocols can be described as following:

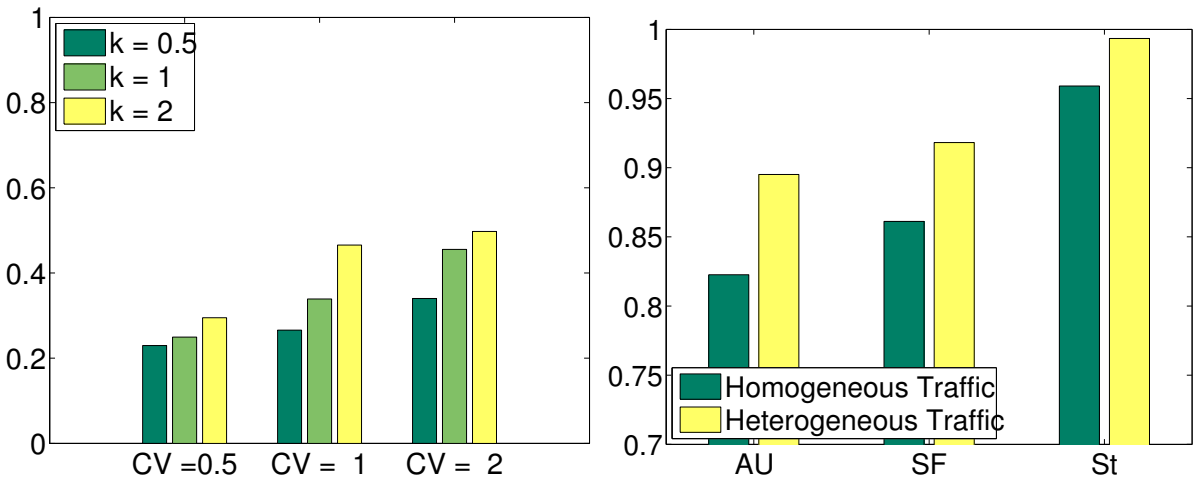
Definition 10. *The probability p_i a node i to be selected as a relay for a message destined to node d, is related to their contact rate λ_{id} and this relation is described by a function $p(\lambda_{id})$.*

As an example, we present two protocols belonging to the above class and their $p(\lambda)$ functions: (a) a modified mobility-aware version of *spray and wait* [129] protocol (we refer to it as *U1*), and (b) a variation of the *EBR* [98] protocol (we refer to it as *U2*), where each relay can hold only one message copy.

U1: A node i, which would be selected as a relay by the *spray and wait* mechanism, under *U1* becomes a relay with a probability p_i that is proportional to its contact rate with the destination d, i.e. $p_i = p(\lambda_{id}) = c \cdot \lambda_{id}$, where c a normalizing factor such as $p(\lambda) \in [0, 1]$.

U2: For each message copy, the source node s selects the relay node i with a probability p_i that is computed according to the *EBR* mechanism, i.e. $p_i = p(\lambda_{id}) = \frac{\lambda_{id}}{\lambda_{id}+\lambda_{sd}}$.

In the following corollary, we prove that our Results 9 and 10 can be simply modified and capture such mobility-aware protocols as well. Corollary 3 follows from a similar analysis as in Section 4.3, whose main analytical steps are described in Section 4.8.2.



(a) Synthetic scenarios (b) Real networks

Figure 4.6: $P_{(src.)}$ of mobility-aware routing in (a) synthetic scenarios with varying mobility (CV_λ) and traffic heterogeneity (k), and (b) real networks with homogeneous and heterogeneous traffic.

Corollary 3. *Under a mobility-aware Relay-Assisted protocol conforming to Def. 10, Results 9 and 10 are modified as:*
Result 9: f_R is given by the L-fold convolution of $f_u(\lambda)$, where

$$f_u(x) = \frac{1}{E[p(\lambda)]} \cdot p(x) \cdot f_\lambda(x)$$

Result 10: The number of copies L is multiplied by c_u, where

$$c_u = \frac{E[\lambda \cdot p(\lambda)]}{E[\lambda] \cdot E[p(\lambda)]}$$

For instance, applying Corollary 3, the expression for the delay ratio R, becomes

$$1 \geq \quad R \quad \geq \quad R_{min} = \frac{1 + (k-1) \cdot CV_\lambda^2}{1 + k \cdot CV_\lambda^2 + c_u \cdot L} \tag{4.23}$$

and for the *U1* protocol presented above, c_u is given by the expression[15]:

$$c_u^{(U1)} = 1 + CV_\lambda^2 \tag{4.24}$$

When mobility is highly heterogeneous (i.e. high CV_λ), $c_u^{(U1)}$ becomes large, and thus R_{min} and P_{min} decrease compared to the random replication mechanism (e.g. random SnW). This confirms that the performance gain is larger when mobility-aware protocols are used. However, even in this case, as traffic heterogeneity increases, the performance gain diminishes, i.e. $R_{min}, P_{min} \to 1$.

We further demonstrate some preliminary simulation results suggesting that our conclusions hold also for mobility-aware routing. We use the *U2* protocol presented above. In Fig. 4.6(a)

[15] An expression for c_u in the case of the *U2* protocol could also be derived, albeit with more complexity, due to the fact that the function $p(\lambda)$ involves the source destination contact rate λ_{sd} as well.

we present simulation results for the delivery metric $P_{(src.)}$ on synthetic scenarios with varying mobility (CV_λ) and traffic (k) heterogeneity. Similarly to the random replication case, for increasing heterogeneity (in mobility and/or traffic) the gain of the extra copies clearly decreases (i.e. $P_{(src.)}$ increases) under mobility-aware schemes. In Fig. 4.6(b) we compare the probability $P_{(src.)}$ of scenarios with and without traffic heterogeneity in real networks. As before, the results are consistent with our theory: the gain of extra copies decreases even for protocols using more sophisticated techniques for relay selection.

Routing based on Contact Graph Structure

A number of mobility-aware routing schemes, e.g. SimBet [29], BubbleRap [60], are based on the structure of the contact graph (centrality, similarity, communities, etc.) rather than the pairwise contact rates. A direct mapping to a function $p(\lambda)$ (Def. 10) for these protocols requires a separate, and rather cumbersome analysis for each such protocol, in most cases not leading to a closed form expression (see, e.g. [112]). However, the contact graphs used to make forwarding decisions by these more sophisticated protocols, still are built based on pair-wise contact rates [52]. We thus expect the utility of such mobility-related information to be similarly affected by the amount of traffic heterogeneity and its relation to mobility patterns.

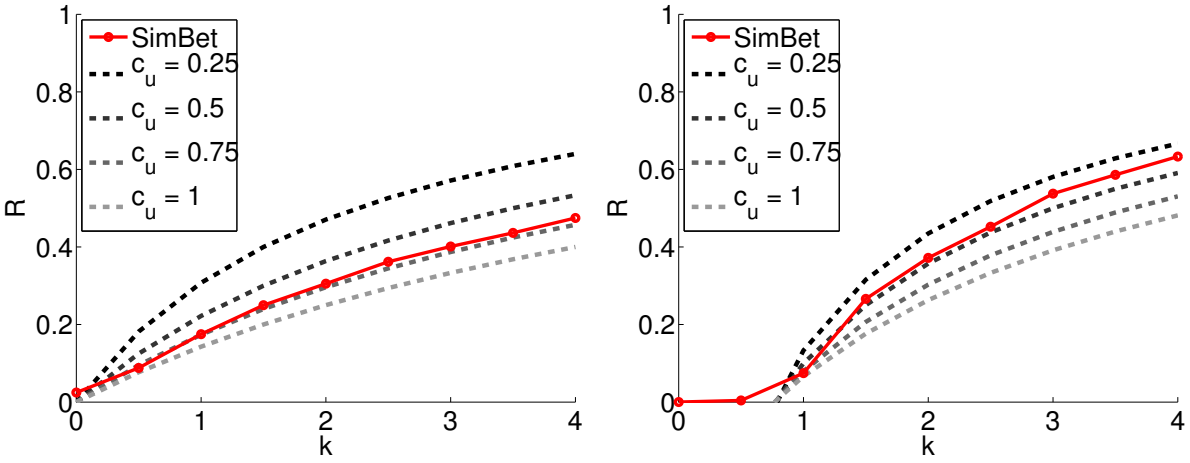
To test this further, we simulated, as an example, scenarios using the SimBet protocol [29][16]. In Fig. 4.7 we present the simulation results (continuous lines) for the ratio $R = \frac{E[T_{SimBet}]}{E[T_{DT}]}$ and the theoretical predictions R_{min} of Eq. (4.23), for different values of the c_u parameter (dashed lines). Two main observations that confirm our intuition are: (i) simulated and theoretical curves increase in a similar manner, and (ii) one can find (numerically) the value c_u that more accurately predicts the performance.

Although this is clearly not conclusive for the applicability of our result to every mobility-aware scheme, we believe it helps to corroborate our findings for the interplay between mobility and traffic heterogeneity on protocol performance.

4.5.2 Multicast Communication

We have also been assuming unicast messages between a $\{s, d\}$ pair. However, our results apply also to multicast [36] or anycast (e.g. content sharing or service composition applications) [119] messages from s, with d being one of the destinations, since similar mechanism are often used for their dissemination. To demonstrate this, in Table 4.6 we present simulation results for two multicast scenarios, with homogeneous (HOM) and heterogeneous (HET) traffic ($\tau(x) = c \cdot x^4$), under varying mobility heterogeneity. A source sends messages to 5 destinations (each selected with a probability $\propto \tau_{ij}$) either by Direct Transmission or by Relay-Assisted routing with $L = 5$ copies. As delivery delay, we consider the delay till all the destinations get the message. It is evident that R and $P_{(src.)}$ (i) increase significantly with mobility heterogeneity when traffic is heterogeneous, and (ii) become much larger compared to the homogeneous case (where R decreases and $P_{(src.)}$ is constant), which is in agreement with our results.

[16] For the contact graph we considered the 10% most frequently meeting pairs following the guidelines of [52], we set the similarity and betweenness weights $\alpha = \beta = 0.5$ [29], and we generated multiple copies as in [30].



(a) $CV_\lambda = 1$ (b) $CV_\lambda = 2$

Figure 4.7: Delay ratio R in two scenarios with varying traffic heterogeneity k. Relay-assisted routing is SimBet with (a) $L = 5$ and (b) $L = 10$ message copies.

Table 4.6: Multicast Communication

	CV_λ	0.1	0.5	1	1.5	2
HOM	R	0.18	0.12	0.01	0	0
	$P_{(src.)}$	0.01	0.01	0.01	0.01	0.01
HET	R	0.18	0.26	0.39	0.52	0.61
	$P_{(src.)}$	0.01	0.03	0.12	0.26	0.41

4.6 Related Work

Useful implications for mobile social networking have arisen from the investigation of *mobility/social ties* and *social ties/communication traffic* correlations, which have been studied extensively and under different disciplines, like anthropology [41], sociology [31], social media [37] or pervasive social networks [109]. For example, [37] shown that the amount of exchanged communication traffic between users of OSNs depends on their social relationships.

On the other hand, the *communication traffic / mobility* correlation has not been given similar attention. There exist only a few works [54, 138] studying it in a framework relevant to MSNs. In [54], Hossmann *et al.* collected and analysed two datasets from online social networks (Facebook and Gowalla / Twitter), and investigated the relations among three dimensions: *mobility, social ties, communication traffic*. With respect to our study, they found that there is strong dependence between mobility and traffic, and, specifically, node pairs that contact during the experiments' duration, communicate with higher probability than the other pairs. Correspondingly, authors in [138] analysed a massive dataset of Call Detail Records (CDRs) of 6 million users and shown a positive correlation between the mobility and communication traffic patterns. Not only they shown that the higher the contact rate (λ_{ij}) of a node pair is, the higher the probability that the nodes communicate intensively, but also found that information inferred by the mobility patterns can work as a good predictor for future communication events. However, despite the fact that [54,138] show clearly that communication traffic is heterogeneous (and correlated to mobility), to our best knowledge, its *effects* on communication performance have not been studied previously.

Finally, with respect to our results and the insights obtained from them, it has already been observed [59, 92] that realistic mobility patterns (e.g. locality, community) can hurt the performance of Relay-Assisted routing (especially simple, random protocols [129]). However, this is a performance degradation that is due to the relays being *too similar* to the source (e.g. all in the same community [59] or with common characteristics [92]). Instead, the *relative* performance degradation here comes due to the source and relays being *too different* in terms of their encounter rates with the destination.

4.7 Conclusions

Motivated by (i) recent findings indicating heterogeneous traffic patterns in mobile social networks and (ii) the lack of related studies, we modelled traffic heterogeneity and studied how it affects the performance in MSNs. We found that the effects can be significant, changing our understanding of common design principles, such as the added value of relays. Despite the fact that some of our qualitative conclusions seem to be rather intuitive, they have not attracted any focus in previous studies, where performance analysis of communication schemes is conducted assuming homogeneous traffic. This indicates a necessity for revisiting the evaluation of protocols in scenarios that entail diversity in the traffic exchanged between nodes.

We believe that our study provides an initial understanding on the effects of traffic heterogeneity. However, traffic patterns in real networks might have much more complex characteristics than what can be captured by our framework, e.g. time-dependent traffic/mobility correlations. Therefore, for a more complete characterisation of traffic demands in mobile social networking (either for end-to-end or content-centric applications [50, 119]), we believe that further experimental (e.g. measurements, recognition of traffic patterns in available datasets, etc.) and

analytical research is needed.

Moreover, our results have some interesting implications about the usefulness of MSNs for various applications. Specifically, common communication traffic patterns seem to have a positive impact on the performance of content-centric applications, rendering them a promising direction for future MSNs. To this end, in the following chapters, we turn our attention to content-centric opportunistic communication, and investigate various aspects related to mobility/traffic patterns and communication performance.

4.8 Appendix: Supplementary Theoretical Results and Proofs

4.8.1 Mobility Independent Heterogeneous Traffic

Heterogeneous communication traffic patterns that are independent of the underlying mobility, can be captured by the following definition (with respect to Def. 9).

Definition 11 (Mobility Independent Heterogeneous Traffic)**.** *The end-to-end traffic demand (per time unit) between a pair of nodes $\{i, j\}$, is a random variable τ_{ij}, with finite mean value $E[\tau_{ij}] = \mu_\tau$, $\mu_\tau \in (0, \infty)$.*

Then, under Def. 11, Lemma 5 states that the effective contact rate between sources and destinations ($\lambda_{sd} \sim f_\tau(\lambda)$) is not different than the contact rate between a randomly chosen pair of nodes ($\lambda_{ij} \sim f_\lambda(\lambda)$). Therefore, it follows evidently that neither the *average* communication performance will be affected by traffic heterogeneity, when it is mobility independent.

Lemma 5. *The probability density function f_τ of the contact rate between the source and the destination $\{s, d\}$ of a random message, in a network following Def. 3 and 11, converges in probability as follows:*

$$f_\tau(x) \xrightarrow{p} f_\lambda(x)$$

Proof. Let us consider the same notation and methodology as in the proof of Proposition 1. The key difference is that now (under Def. 11), the mean value of the random variables τ_{ij} is μ_τ (i.e. independent of mobility). Thus, it holds that for a large network (*weak law of large numbers*)

$$X_1 = \frac{T(\lambda)}{\|\mathcal{N}(\lambda)\|} \xrightarrow{p} \mu_\tau \quad \text{and} \quad X_3 = \frac{\sum_i \sum_i \tau_{ij}}{N(N-1)/2} \xrightarrow{p} \mu_\tau$$

Then, applying Cramér's theorem, gives

$$X_1 \cdot X_2 \cdot \frac{1}{X_3} \xrightarrow{p} \mu_\tau \cdot f_\lambda(\lambda) \cdot \frac{1}{\mu_\tau} = f_\lambda(\lambda)$$

which proves the Lemma. □

4.8.2 Mobility Aware Protocols - Proof of Corollary 3

Since the relay selection is mobility dependent, the contact rates between relays and destinations will *not* be distributed with f_λ. Following similar arguments as in the proof of Proposition 1, it can be shown that for mobility-aware protocols that follow the model presented in Section 4.5, Lemma 6 holds.

Lemma 6. *The contact rate between a relay and the destination of a random message, under mobility-aware routing, converges in probability as follows:*

$$f_u(x) \xrightarrow{p} \frac{1}{E[p(\lambda)]} \cdot p(x) \cdot f_\lambda(x) \tag{4.25}$$

where $E[p(\lambda)] = \int_0^\infty p(x) f_\lambda(x) dx$

Using Lemma 6, Results 9 and 10 are modified as following:

Result 9: The function f_R used in Result 9 (see also Eq. (4.6)-Eq. (4.7) in its proof) will be now the L-fold convolution of $f_u(\lambda)$ (rather than f_λ), i.e. $f_R = f_u^{(*L)}$.

Result 10: Since $f_R = f_u^{(*L)}$, the mean value $E_R[y]$ (see Eq. (4.12)) will be given now by

$$E_R[y] = L \cdot E_u[y] = L \cdot \int_0^\infty y \cdot f_u(y) dy \tag{4.26}$$

Substituting f_u from Eq. (4.25) to Eq. (4.26), gives

$$E_R[y] = L \cdot \int_0^\infty y \cdot \frac{p(y)}{E[p(\lambda)]} \cdot f_\lambda(y) dy = L \cdot \frac{E[\lambda \cdot p(\lambda)]}{E[p(\lambda)]} \tag{4.27}$$

where the expectations are taken over f_λ.

Comparing Eq. (4.12) and Eq. (4.26), it is easy to see that one need to replace the term $L \cdot \mu_\lambda$ with a term $L \cdot \frac{E[\lambda \cdot p(\lambda)]}{E[p(\lambda)]} = c_u \cdot L \cdot \mu_\lambda$, where

$$c_u = \frac{E[\lambda \cdot p(\lambda)]}{\mu_\lambda \cdot E[p(\lambda)]} = \frac{E[\lambda \cdot p(\lambda)]}{E[\lambda] \cdot E[p(\lambda)]} \tag{4.28}$$

Chapter 5

Content-Centric Traffic: Effect of Content Popularity and Availability Patterns

5.1 Introduction

The proliferation of "smart" mobile devices has led researchers to consider MSNs as a way to support existing infrastructure and/or novel applications, like file sharing [34, 142], crowd sensing [105, 123], collaborative computing [26, 119], offloading of cellular networks [50, 85, 141], etc. While MSNs initially proposed for end-to-end communications, lately the trend is shifting to *content-centric* communications. Some content-centric applications for which mobile social networking has been considered are: (i) *content sharing* [10, 34, 99]: the source(s) of a "content" (e.g. multimedia file, web page) might want to distribute it (e.g. user generated content) or is willing to share it with other nodes (e.g. content downloaded earlier); (ii) *service or resource access* [26, 119]: nodes offer access to resources (e.g. Internet access) or services (e.g. computing resources); (iii) *mobile data offloading* [50, 85, 141]: the cellular network provider, instead of serving separately each node requesting a given "content" (e.g. a popular video, or software update), distributes a few copies of the "content" in some relay nodes (or *holders*) and they can further forward it to any other node that makes a request for it.

The performance of these mechanisms highly depends on *who* is interested, in *what*, and *where* it can be found (i.e. which other nodes have it). While the effect of node mobility has been extensively considered (e.g. [10, 34, 113]) content popularity has been mainly considered from an algorithmic perspective (e.g [85, 99]), and in the context of a specific application. Despite the inherent interest of these studies, some questions remain:

Would a given allocation policy work well in a different network setting? Are there interest patterns that would make a scheme generally better than others?

Key factors like content popularity and content availability might impact the performance or even decide the feasibility of a given application altogether. In this paper, we try to provide some initial insight into these questions, by contributing along the following key directions:

- We propose a simple analytical framework that is applicable to a range of mobility and content popularity patterns seen in real networks; to our best knowledge, this is the first application-independent effort in this direction (Section 5.2).

- We provide closed form expressions for important metrics that require few statistics about the aggregate node mobility and content popularity; these results facilitate *online* performance prediction and protocol tuning, compared to approaches requiring detailed per node statistics, as e.g. [85] (Section 5.3).
- While a detailed application-specific optimization is beyond the scope of this paper, we demonstrate how our analysis can be applied to an example application, mobile data offloading, and can help optimize its performance in a generic setting (Section 5.4).

5.2 Content-Centric Traffic Model

We consider a network $\mathcal{N}$ with N nodes moving according to the model of Def. 3[1]

We assume that each node might be *interested in* one or more "contents". A content of interest might refer to (i) a single piece of data (e.g. a multimedia file, a google map) [141], (ii) all messages/data belonging to a category of interests (e.g. local events, financial news) [28,142], (iii) updates and feeds (e.g. weather forecast, latest news) [79], etc.

A number of content-sharing applications and mechanisms have been proposed in previous literature, from publish-subscribe mechanisms to "channel"-based sharing and device-to-device offloading, etc., (e.g. [79,105,123,142]). To proceed with our analysis we need to setup a simple model of content/service access that can yet capture different (but of course not all) content-centric applications and approaches.

The main notation we use in our model and analysis is summarized in Table 5.1.

5.2.1 Content Popularity

We assume that when a node is interested in a content or service, it queries other nodes it *directly* encounters for it. We denote the event that a node $i \in \mathcal{N}$ is *interested in* a content $\mathcal{M}$ (or, equivalently, i requests $\mathcal{M}$) as: $i \to \mathcal{M}$. We further denote the set of all the contents that nodes are interested in, as: $\mathbf{M} = \{\mathcal{M} : \exists i \in \mathcal{N}, i \to \mathcal{M}\}$. $|\mathbf{M}| = M$, where $|\cdot|$ denotes the cardinality of a set.

Definition 12 (Content Popularity)**.** *We define the popularity of a content $\mathcal{M}$ as the number of nodes $N_p^{(\mathcal{M})}$ that are interested in it*[2]*:*

$$N_p^{(\mathcal{M})} = |\mathcal{C}_p^{(\mathcal{M})}|, \text{ where } \mathcal{C}_p^{(\mathcal{M})} = \{i \in \mathcal{N} : i \to \mathcal{M}\} \tag{5.1}$$

We further denote the percentage of contents with a given popularity value n as

$$P_p(n) = \frac{1}{M} \sum_{\mathcal{M} \in \boldsymbol{M}} \mathcal{I}_{N_p^{(\mathcal{M})}=n}, \quad n \in [0, N] \tag{5.2}$$

where $\mathcal{I}_{N_p^{(\mathcal{M})}=n} = 1$ when $N_p^{(\mathcal{M})} = n$ and 0 otherwise.

In other words, $P_p(n)$ defines a probability distribution over the different contents and associated popularities. In practice, it can be chosen according to common practices (e.g. skewed, pareto) [34,85,99], or be fitted to real data, if available.

[1] Our main results can be extended for a contact network, similar to this of Def. 3, but where inter-contact intervals are Pareto distributed (see Section 5.3.3.3).

[2] This could be an average, calculated over some time window.

5.2.2 Content Availability

We assume that a request for a content or service is completed, when (and if) a node that holds (a copy of) the requested content is *directly* encountered. We denote the event that a node i holds (a copy of) a content $\mathcal{M}$ as $i \leftarrow \mathcal{M}$, and we define the availability $N_a^{(\mathcal{M})}$ of a content $\mathcal{M}$ as

Definition 13 (Content Availability)**.** *The availability of a content message $\mathcal{M}$ is defined as the number of nodes $N_a^{(\mathcal{M})}$ that hold a copy of it.*

$$N_a^{(\mathcal{M})} = |\mathcal{C}_a^{(\mathcal{M})}|, \text{ where } \mathcal{C}_a^{(\mathcal{M})} = \{i \in \mathcal{N} : i \leftarrow \mathcal{M}\} \tag{5.3}$$

The availability of a given content might often (although not always) be correlated with the popularity of that content. A cellular network provider might *allocate* more holders for popular contents [85]. In a content-sharing setting, where some nodes might be more willing than others to maintain and share ("seed") a content after they've downloaded and "consumed" it, popular content will end up being shared by more nodes. We will model such correlations in a *probabilistic* way, as follows.

Definition 14 (Availability vs. Popularity)**.** *The availability of* any *content message $\mathcal{M}$ is related to its popularity through the relation*

$$P\{N_a^{(\mathcal{M})} = m | N_p^{(\mathcal{M})} = n\} = g(m|n) \tag{5.4}$$

The above conditional probabilities can describe a wide range of cases where availability depends on popularity, and some additional randomness might be present due to factors like: natural churn in the nodes sharing the content, content-dependent differences in the sharing policies applied by nodes, estimation noise, etc. Some special cases of this model include: (i) *uncorrelated availability*, where $g(m|n) \equiv g(m)$; (ii) *deterministic availability*, where:

$$N_a^{(\mathcal{M})} = \rho\left(N_p^{(\mathcal{M})}\right) \quad \Leftrightarrow \quad g(m|n) = \begin{cases} 1, & m = \rho(n) \\ 0, & \text{otherwise} \end{cases}$$

where $\rho(n) : [1, N] \rightarrow [0, N]$ can be an arbitrary function. One such example could be a deterministic approximation of $g(m|n)$ with its average value, namely $\rho(n) = \bar{g}(n) \equiv \sum_m m \cdot g(m|n)$.

5.3 Analysis of Content Requests

We will now analyze how different popularity, availability, and mobility patterns (possibly arising from different applications, policies, and network settings) affect key metrics like: (i) the delay to access a content of interest, (ii) the probability to retrieve a content before a deadline. A key parameter for these metrics is the number of holders for the requested content (availability). The higher this number, the sooner a requesting node will encounter one of them.

While content availability might sometimes be time dependent [99], or the content holders might be chosen based on their mobility properties [85], we first make two additional assumptions that allow us to derive simple, useful expressions. In Section 5.3.3, we relax both these assumptions.

Table 5.1: Important Notation

CONTENT TRAFFIC (Section 5.2)		
$i \to \mathcal{M}$	Node i *is interested / requests* content $\mathcal{M}$	
$\mathbf{M}$	Set of contents in the network, $\lvert\mathbf{M}\rvert = M$.	
$N_p^{(\mathcal{M})}$	Popularity of content $\mathcal{M}$	Def. 12
$\mathcal{C}_p^{(\mathcal{M})}$	Set of nodes interested in content $\mathcal{M}$	Def. 12
$P_p(n)$	Probability distribution of content popularity	Eq. (5.2)
$i \leftarrow \mathcal{M}$	Node i *holds a copy* of content $\mathcal{M}$	
$N_a^{(\mathcal{M})}$	Availability of content $\mathcal{M}$	Def. 13
$\mathcal{C}_a^{(\mathcal{M})}$	Set of nodes that hold a copy of content $\mathcal{M}$	Def. 13
$g(m\|n)$	Availability - Popularity relation	Def. 14
$\rho(n)$	Deterministic case for $g(m\|n)$	
$\overline{g}(n)$	The average value of $g(\cdot\|n)$	
ANALYSIS (Section 5.3.1)		
$P_p^{req.}(n)$	Popularity distribution of a random request	Lemma 7
$P_a^{req.}(n)$	Availability distribution of a random request	Lemma 8
T_{ij}	Time of *next* meeting between nodes i and j	
$T_{\mathcal{M}}$	Content access time	
$X_{\mathcal{M}}$	Sum of meeting rates of j and nodes $\in \mathcal{C}_a^{(\mathcal{M})}$	Eq. (5.6)

Assumption 1. *The popularity* $N_p^{(\mathcal{M})}$ *and availability* $N_a^{(\mathcal{M})}$ *of a content* $\mathcal{M}$ *do not change over time.*

Assumption 2. *The set of requesters* $\mathcal{C}_p^{(\mathcal{M})}$ *and holders* $\mathcal{C}_a^{(\mathcal{M})}$ *of a content* $\mathcal{M}$ *are independent of node mobility.*

Assumption 1 is valid (or a good approximation), for example, when the number of holders is chosen by the cellular operator [50,85] or content provider, and other nodes cannot act as holders or do not have incentives to do so. It is also valid when a given service (e.g. Internet access, or specific sensor) is offered only by a certain number of devices [119], or the "content" refers to a *channel* or *category* and not a particular file [79]. Nevertheless, if a content is disseminating and new nodes are willing to share it [141], then its availability might change over time.

Assumption 2 is a reasonable approximation when a *mobility oblivious* allocation policy is considered (e.g. [99], or the homogeneous algorithm of [85]) or when there is no knowledge of the interests-mobility correlation, if any. Nevertheless, there exist scenarios where who holds what content might depend on the contact rates with other nodes [10, 85].

5.3.1 Preliminary Analysis

Assume we observe the network for a long time, during which a large number of requests have been made. Assume further that we pick one such request randomly, which happens to be for content $\mathcal{M}$, and we want to predict its performance. We need to answer the following two questions:

Q.1 What is the popularity of $\mathcal{M}$?

Q.2 How fast does a requesting node meet $\mathcal{M}$'s holders?

Q.1 is needed to predict the availability for $\mathcal{M}$, which according to Assumption 1 does not depend on the exact time of the request. Given the availability of $\mathcal{M}$, Q.2 will estimate the (sum of) contact rates between the requesting node and the holders, according to Assumption 2 and the mobility model.

Answering **Q.1**

It is easy to see that the popularity of $\mathcal{M}$ should be proportional to $P_p(n)$: the higher the number of different contents with a popularity value n, the higher the chance that $\mathcal{M}$ will be of popularity n. However, the higher the popularity of a content, the more the requests made for it. Hence, a first important observation is that the popularity of the content of such a *random request* is *not* distributed as $P_p(n)$ but is also proportional to the popularity value n.

Consider a stylized example, where only two contents exist in the network, content A with popularity value 10 and content B with popularity value 1. Hence, "half" the contents are of high popularity (10), and "half" of low (1), or in other words $P_p(10) = P_p(1) = \frac{1}{2}$. However, if we observe the network for a long time, 10 times more requests will be made, on average, for content A. Consequently, if we select a request randomly, there is a 10× higher chance that it will be for content A, that is, for the content of popularity 10. Normalizing to have a proper probability distribution gives us the following lemma.

Lemma 7. *The probability that a random request is for a content of popularity equal to n is given by*

$$P_p^{req.}(n) = \frac{n}{E_p[n]} \cdot P_p(n)$$

where $E_p[n] = \sum_n n \cdot P_p(n)$ is the average content popularity [3].

Answering **Q.2**

The answer to question Q.2 consists of two separate steps: (i) we calculate the number of holders of $\mathcal{M}$, and then (ii) we calculate how fast the requesting node can meet these holders. Towards answering (i), Lemma 8 maps the popularity of the content involved in a random request (derived in Lemma 7) to the number of holders for this content. This number is a random variable dependent both on the popularity distribution $P_p(n)$, and on the availability function $g(m|n)$.

Lemma 8. *The probability that a random request is for a content of availability equal to m is given by*

$$P_a^{req.}(m) = \frac{E_p[n \cdot g(m|n)]}{E_p[n]}$$

Proof. For a random request for content $\mathcal{M}$, using the property of conditional expectation, we can write [121]:

$$P_a^{req.}(m) = \sum_n P\{N_a^{(\mathcal{M})} = m | N_p^{(\mathcal{M})} = n\} \cdot P_p^{req.}(n)$$

where $P_p^{req.}(n)$ is defined in Lemma 7. Substituting, from Def. 14 and Lemma 7, the above terms, we successively get

$$P_a^{req.}(m) = \sum_n g(m|n) \cdot \frac{n}{E_p[n]} \cdot P_p(n) = \frac{\sum_n g(m|n) \cdot n \cdot P_p(n)}{E_p[n]} = \frac{E_p[n \cdot g(m|n)]}{E_p[n]}$$

which proves the Lemma. □

Having computed the statistics for the content availability, we can now calculate how fast the requesting node, say j, meets *any* of the holders i (i.e. nodes $i \in \mathcal{C}_a^{(\mathcal{M})}$). As defined in Def. 3, the inter-contact intervals are exponentially distributed. Hence, let T_{ij} denote the inter-contact times between node j and a node $i \in \mathcal{C}_a^{(\mathcal{M})}$, and let T_{ij} be exponentially distributed with rate λ_{ij}. If we denote with $T_{\mathcal{M}}$ the first time until j meets *any* of the nodes $i \in \mathcal{C}_a^{(\mathcal{M})}$ (and, thus, accesses the content), then: $T_{\mathcal{M}} = \min_{i \in \mathcal{C}_a^{(\mathcal{M})}} \{T_{ij}\}$, i.e. $T_{\mathcal{M}}$ is distributed as a minimum of exponential random variables, and it holds that [121]:

$$T_{\mathcal{M}} \sim exp\,(X_{\mathcal{M}}) \quad \Leftrightarrow \quad P\{T_{\mathcal{M}} > t\} = e^{-X_{\mathcal{M}} \cdot t} \tag{5.5}$$

[3] We use subscript p to denote an expectation over the popularity distribution $P_p(n)$, and n denotes the random popularity values.

Table 5.2: Performance Metrics when $f_\lambda \sim Gamma$ with μ_λ, CV_λ and $P_p(n) \sim Pareto(n_0, \alpha = 2)$.

$\rho(n) = c \cdot n$	$E[T_\mathcal{M}] = \frac{1}{\mu_\lambda \cdot CV_\lambda^2} \left[\frac{c \cdot n_0}{CV_\lambda^2} \cdot \ln \left(\frac{1}{1 - \frac{CV_\lambda^2}{c \cdot n_0}} \right) - 1 \right]$
$\rho(n) = c \cdot \ln(n)$	$P\{T_\mathcal{M} \leq TTL\} = 1 - \frac{1}{(1+\ln(\gamma)) \cdot \gamma^{\ln(n_0)}}$ where $\gamma = (1 + \mu_\lambda \cdot CV_\lambda^2 \cdot TTL)^{\frac{c}{CV_\lambda^2}}$

where

$$X_\mathcal{M} = \sum_{i \in \mathcal{C}_a^{(\mathcal{M})}} \lambda_{ij} \tag{5.6}$$

Clearly, knowing $X_\mathcal{M}$ is needed to proceed with the desired metric derivation. Based on the preceding discussion, $X_\mathcal{M}$ is a random variable that depends on: (i) the number of content holders m (i.e. the cardinality of set $\mathcal{C}_a^{(\mathcal{M})}$ in Eq.(5.6)), and (ii) the meeting rates with the holders. Applying Assumption 2, it holds that, conditioning on m, $X_\mathcal{M}$ (Eq. (5.6)) is a sum of m i.i.d. random variables $\lambda_{ij} \sim f_\lambda(\lambda)$, i.e

$$X_\mathcal{M} \sim f_{m\lambda}(x) = (f_\lambda * f_\lambda \cdots * f_\lambda)_m \,, \tag{5.7}$$

where $*$ denotes convolution, and mean value [121]:

$$E[X_\mathcal{M} | N_a^{(\mathcal{M})} = m] = E_{m\lambda}[x] = m \cdot \mu_\lambda \tag{5.8}$$

5.3.2 Performance Metrics

We consider two main performance metrics: the *average delay* and *delivery probability*. Based on the analysis of Section 5.3.1, we derive results under generic content traffic (i.e. $P_p(n)$ and $g(m|n)$) and mobility (i.e. $f_\lambda(\lambda)$) patterns.

5.3.2.1 Content Access Delay

Result 11. *The expected content access delay can be computed with the expression*

$$E[T_\mathcal{M}] = \frac{1}{E_p[n]} \cdot E_p \left[n \cdot \sum_m E_{m\lambda} \left[\frac{1}{x} \right] \cdot g(m|n) \right]$$

Proof. The time $T_\mathcal{M}$ a node j needs to access a content $\mathcal{M}$ is exponentially distributed with rate $X_\mathcal{M}$. However, $X_\mathcal{M}$ is a random variable itself, distributed with $f_{m\lambda}(x)$ (Eq. (5.7)). Thus, we can write for the expected content access delay:

$$\begin{aligned} E[T_\mathcal{M}] &= \sum_m E[T_\mathcal{M} | N_a^{(\mathcal{M})} = m] \cdot P_a^{req.}(m) \\ &= \sum_m \int E[T_\mathcal{M} | X_\mathcal{M} = x, N_a^{(\mathcal{M})} = m] \cdot f_{m\lambda}(x) dx \cdot P_a^{req.}(m) = \sum_m \int \frac{1}{x} \cdot f_{m\lambda}(x) dx \cdot P_a^{req.}(m) \end{aligned} \tag{5.9}$$

The last equality follows from the fact that the expectation of an exponential random variable with rate x is $\frac{1}{x}$.

Expressing the integral in Eq. (5.9) as an expectation over the $f_{m\lambda}(x)$ and substituting $P_a^{req.}(m)$ from Lemma 8, gives

$$E[T_{\mathcal{M}}] = \sum_m E_{m\lambda}\left[\frac{1}{x}\right] \cdot \frac{E_p[n \cdot g(m|n)]}{E_p[n]} = \frac{1}{E_p[n]} \cdot \sum_m E_{m\lambda}\left[\frac{1}{x}\right] \cdot E_p[n \cdot g(m|n)] \tag{5.10}$$

Rearranging the expectations and summation in Eq. (5.10) we get the expression of Result 11. □

If the functions $f_\lambda(\lambda)$, $g(m|n)$ and $P_p(n)$ are known, the expected delay $E[T_{\mathcal{M}}]$ can be computed directly from Result 11, as shown in the following example.

Example Scenario: The contact rates (f_λ) follow a *gamma distribution*, as suggested in [107], with μ_λ and CV_λ. Content popularity $P_p(n)$ is Pareto distributed, as observed in [83], with *scale* and *shape* parameters n_0 and $\alpha = 2$, respectively. Finally, we consider a (deterministic) allocation of holders, $\rho(n) = c \cdot n$ (see Section 5.2.2). Then a closed form expression for $E[T_M]$ is given in the first row of Table 5.2.

However, in a real implementation, it might not be always possible to know the *exact* distributions of the contact rates (f_λ) and/or the availabilities ($g(m|n)$), needed to compute the expression of Result 11. In the following theorem, we derive an expression for $E[T_M]$ that requires only the *average statistics* (which are much easier to estimate or measure in a real scenario), namely (i) the mean value of the contact rates, μ_λ, and (ii) the average availability for contents of a given popularity, $\overline{g}(n)$.

Theorem 2. *A lower bound for the expected content access delay is given by*

$$E[T_{\mathcal{M}}] \geq \frac{1}{\mu_\lambda \cdot E_p[n]} \cdot E_p\left[\frac{n}{\overline{g}(n)}\right]$$

Proof. In Result 11 we can express $E_{m\lambda}\left[\frac{1}{x}\right]$ as $E_{m\lambda}[h(x)]$, where $h(x) = \frac{1}{x}$. Since $h(x)$ is a convex function, applying *Jensen's inequality*, i.e. $h\left(E[x]\right) \leq E[h(x)]$, gives

$$E_{m\lambda}\left[\frac{1}{x}\right] \geq \frac{1}{E_{m\lambda}[x]} = \frac{1}{m \cdot \mu_\lambda} \tag{5.11}$$

where, in the equality, we used Eq. (5.8).

Substituting Eq. (5.11) in the expression of Result 11, gives

$$E[T_{\mathcal{M}}] \geq \frac{1}{\mu_\lambda \cdot E_p[n]} \cdot E_p\left[n \cdot \sum_m \frac{1}{m} \cdot g(m|n)\right] \tag{5.12}$$

The sum in Eq. (5.12) is the expectation over $g(\cdot|n)$, i.e.

$$\sum_m \frac{1}{m} \cdot g(m|n) = E_g\left[\frac{1}{m}\right] \tag{5.13}$$

Applying, as before, Jensen's inequality, we get

$$\sum_m \frac{1}{m} \cdot g(m|n) = E_g\left[\frac{1}{m}\right] \geq \frac{1}{E_g[m]} = \frac{1}{\overline{g}(n)} \tag{5.14}$$

where we used for $E_g[m]$ the notation $\overline{g}(n)$.

Combining Eq. (5.14) and Eq. (5.12), the expression of the theorem follows directly. □

5.3.2.2 Content Access Probability

One often needs to also know the probability that a node can access a content by some deadline, i.e. $P\{T_{\mathcal{M}} \leq TTL\}$. E.g, a node might lose its interest in a content (e.g. news) after some time, or in an offloading scenario a node might decide to access a content directly to the base station.

Result 12. *The probability a content to be accessed before a time TTL can be computed with the expression*

$$P\{T_{\mathcal{M}} \leq TTL\} = 1 - \frac{E_p\left[n \cdot \sum_m E_{m\lambda}\left[e^{-x \cdot TTL}\right] g(m|n)\right]}{E_p[n]}$$

Proof. Conditioning on the values of $N_a^{(\mathcal{M})}$ and $X_{\mathcal{M}}$, as in Eq. (5.9), we can write:

$$P\{T_{\mathcal{M}} \leq TTL\} = \sum_m \int P\{T_{\mathcal{M}} \leq TTL | X_{\mathcal{M}} = x, N_a^{(\mathcal{M})} = m\} \cdot f_{m\lambda}(x)dx \cdot P_a^{req.}(m)$$
$$= 1 - \sum_m \int e^{-x \cdot TTL} \cdot f_{m\lambda}(x)dx \cdot P_a^{req.}(m) \quad (5.15)$$

where the last equality follows because $T_{\mathcal{M}}$ is exponentially distributed with rate $X_{\mathcal{M}} = x$. After some similar steps as in Theorem 2, the final result follows. □

The expression of Result 12 for the example scenario of Section 5.3.2.1, with a different allocation function $\rho(n) = c \cdot \ln(n)$, is given in the second row of Table 5.2.

Theorem 3. *An upper bound for the probability to access a content by a time TTL is given by*

$$P\{T_{\mathcal{M}} \leq TTL\} \leq 1 - \frac{1}{E_p[n]} \cdot E_p\left[n \cdot e^{-\overline{g}(n) \cdot \mu_\lambda \cdot TTL}\right]$$

Proof. The bound follows easily by observing that $h(x) = e^{-x \cdot TTL}$ is a convex function, and applying *Jensen's inequality* and the methodology of Theorem 2. □

5.3.3 Extensions

In this section, we study how the results of Section 5.3.2 can be modified, when we remove the Assumptions 1 and 2, or when we consider pareto distributed intercontact intervals. We state here only the main findings and sketches of the proofs; the detailed proofs can be found in the Section 5.7.

5.3.3.1 Popularity / Availability Time Dependence

Let us assume a scenario where, initially, some nodes hold some *content items* (e.g. data files), in which some other nodes are interested. This can be, for example, a content sharing scenario with contents being, e.g., some google maps. When a node interested in a content item, meets a holder and gets the content, it can hold it in its memory and act as a holder too. Specifically, we describe such scenarios as:

Definition 15.
I. When a requester accesses a content, acts as a holder for it.
II. The initial *content popularity and availability patterns are given by $P_p(n)$ and $g(m|n)$.*

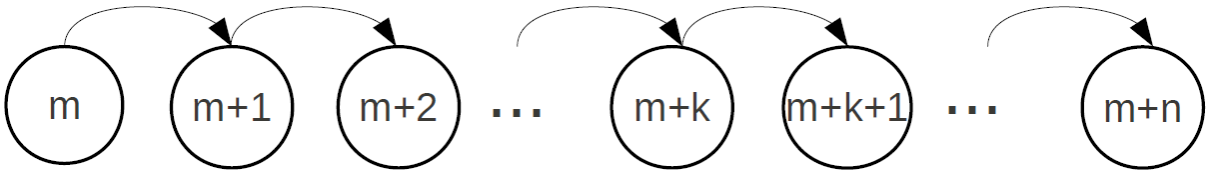


Figure 5.1: Markov Chain for the dissemination of a content with initial popularity and availability n and m, respectively.

Result 13. *Under time changing availability / popularity (Def. 15), the expected content access delay is approximately given by*

$$E[T_{\mathcal{M}}] = \frac{1}{\mu_\lambda \cdot E_p[n]} \cdot E_p\left[\ln\left(1 + \frac{n}{\overline{g}(n)}\right)\right]$$

Sketch of proof: Let us consider a content $\mathcal{M}$ of initial popularity $N_p^{(\mathcal{M})}(0) = n$ and availability $N_a^{(\mathcal{M})}(0) = m$. When the first requester accesses the content, the number of holders will increase to $m+1$ and the remaining requesters will be $n-1$. Building a Markov Chain as in Fig. 5.1, where each state denotes the number of holders, it can be shown for the expected delay of moving from state $m+k$ to state $m+k+1$, $k \in [0,1]$, that it holds $E[T_{k,k+1}] \approx \frac{1}{(m+k)\cdot(n-k)\cdot\mu_\lambda}$. Computing the times $E[T_{k,k+1}]$ and averaging over all the contents, gives the expected delay.

5.3.3.2 Mobility Dependent Allocation

Definition 16 (Mobility Dependent Allocation)**.** *The probability π_{ij} a node i to be a holder for a content in which a node j is interested, is related to their contact rate λ_{ij} such that $\pi_{ij} = \pi(\lambda_{ij})$, where $\pi(\cdot)$ is a function from $\mathbb{R}^+$ to $[0,1]$.*

Result 14. *Under Def. 16, Theorems 2 and 3 and Result 13 hold if we replace μ_λ with $\mu_\lambda^{(\pi)}$, where*

$$\mu_\lambda^{(\pi)} = \frac{E_\lambda[\lambda \cdot \pi(\lambda)]}{E_\lambda[\pi(\lambda)]}$$

where $E_\lambda[\cdot]$ denotes an expectation taken over the contact rates distribution $f_\lambda(\lambda)$.

Sketch of proof: Since the requesters-holders contact rates are mobility dependent, the contact rates between them are not distributed with the contact rates distribution $f_\lambda(\lambda)$, but with a modified version of it, i.e. with a distribution:

$$f_\pi(\lambda) = \frac{1}{E_\lambda[\pi(\lambda)]} \cdot \pi(\lambda) \cdot f_\lambda(\lambda)$$

Hence, Eq. (5.7) and Eq. (5.8) need to be modified as:

$$X_{\mathcal{M}} \sim f_{m\pi}(x) = (f_\pi * f_\pi \cdots * f_\pi)_m$$
$$E[X_{\mathcal{M}} | N_a^{(\mathcal{M})} = m] = E_{m\pi}[x] = m \cdot \frac{E_\lambda[\lambda \cdot \pi(\lambda)]}{E_\lambda[\pi(\lambda)]} = m \cdot \mu_\lambda^{(\pi)}$$

Example Scenario: The holders of a content $\mathcal{M}$ are selected taking into account their contact rates with the requesters, as following: Each node i (candidate to be a holder) is assigned a weight

$w_i = \prod_{j \in \mathcal{C}_p^{(\mathcal{M})}} \lambda_{ij}$. Using such weights, the selection of holders that rarely meet the requesters is avoided. Then, each node is selected to be one of the $N_a^{(\mathcal{M})}$ holders with probability $p_i = \frac{w_i}{\sum_i w_i}$. With respect to Def. 16, it turns out that this mechanism is (approximately) described by $\pi(\lambda) = c \cdot \lambda$. Substituting $\pi(\lambda)$ in Result 14, gives

$$\mu_\lambda^{(\pi)} = \frac{E_\lambda[\lambda \cdot \pi(\lambda)]}{E_\lambda[\pi(\lambda)]} = \frac{E_\lambda[\lambda^2]}{E_\lambda[\lambda]} = \mu_\lambda \cdot (1 + CV_\lambda^2) \tag{5.16}$$

5.3.3.3 Pareto Inter-Contact Times

Although, as discussed in previous chapters, it is common to assume that the times between consecutive contacts for a given pair are exponentially distributed, there exist some studies [19] suggesting a power-law (e.g. pareto) distribution for these intercontact intervals. In this section, we provide the guidelines for applying our analysis to the pareto case.

Let assume that inter-contact times between node j and a node $i \in \mathcal{C}_a^{(\mathcal{M})}$ are pareto distributed with *shape* and *scale* parameters α_{ij} and t_0, respectively:

$$T_{ij} \sim pareto(\alpha_{ij}) \Leftrightarrow P\{T_{ij} > t\} = \left(\tfrac{t_0}{t}\right)^{\alpha_{ij}} \tag{5.17}$$

Then, following the methodology of Section 5.3.1, it can be shown for $T_{\mathcal{M}} = \min_{i \in \mathcal{C}_a^{(\mathcal{M})}}\{T_{ij}\}$ that (Appendix 5.7.3):

$$T_{\mathcal{M}} \sim pareto(A_{\mathcal{M}}) \Leftrightarrow P\{T_{\mathcal{M}} > t\} = \left(\tfrac{t_0}{t}\right)^{A_{\mathcal{M}}} \tag{5.18}$$

where $A_{\mathcal{M}} = \sum_{i \in \mathcal{C}_a^{(\mathcal{M})}} \alpha_{ij}$.

Remark: In this case the contact rates will be $\lambda_{ij} = \frac{1}{E[T_{ij}]} = \frac{1}{t_0} \cdot \left(1 - \frac{1}{\alpha_{ij}}\right)$, $\alpha_{ij} > 1$. However, for simplicity, we can use the parameters α_{ij} instead of the rates λ_{ij}, and, correspondingly, a distribution $f_\alpha(\alpha)$, instead of $f_\lambda(\lambda)$.

Similarly to $X_{\mathcal{M}}$, $A_{\mathcal{M}}$ is a random variable that depends on: (i) the number of content holders m (i.e. the cardinality of set $\mathcal{C}_a^{(\mathcal{M})}$ in Eq.(5.6)), and (ii) the meeting rates with the holders. Hence, the corresponding expressions to Eq. (5.7) and Eq. (5.8) for Pareto intervals ($f_a(\alpha)$, μ_α) are:

$$A_{\mathcal{M}} \sim f_{m\alpha}(x) = (f_\alpha * \cdots * f_\alpha)_m\,, \quad E_{m\alpha}[x] = m \cdot \mu_\alpha$$

Based on the above discussion and the analysis of Section 5.3.2, in Appendix 5.7.4 we derive the expressions for the performance metrics (i.e. expressions corresponding to Results 11 and 12, and Theorems 2 and 3), which we present in Table 5.3.

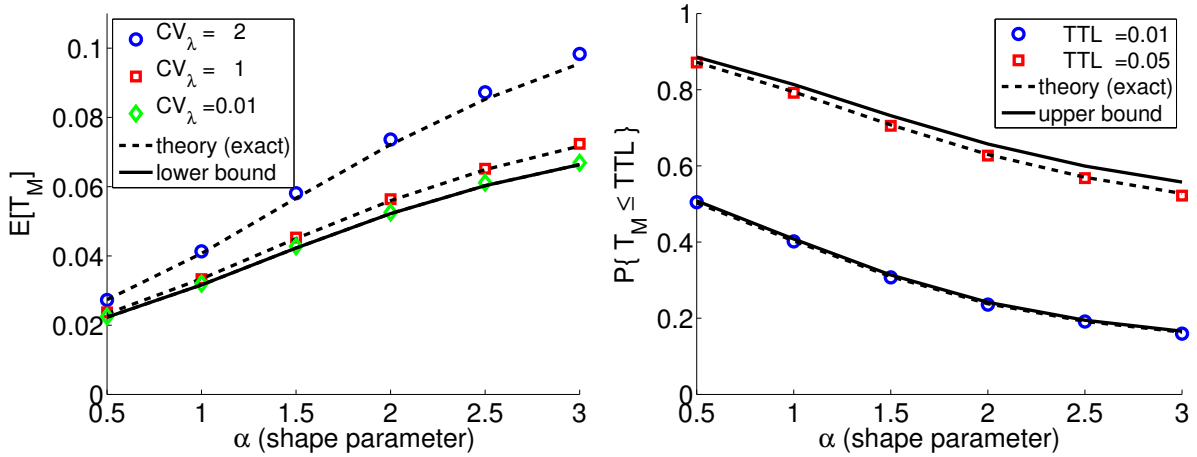
5.3.4 Model Validation

As a first validation step, we compare our theoretical predictions to synthetic simulation scenarios conforming to the models of Section 5.2, in order to consider (a) various mobility and content traffic patterns, and (b) large networks.

Simulation Scenarios: We assign to each pair $\{i, j\}$ a contact rate λ_{ij}, which we draw randomly from a distribution $f_\lambda(\lambda)$, and create a sequence of contact events (Poisson process

Table 5.3: Performance metrics for Pareto distributed Inter-Contact times

	Exact expressions	Bounds
$E[T_{\mathcal{M}}]$	$t_0 + \frac{t_0}{E_p[n]} \cdot E_p\left[n \cdot \sum_m E_{m\alpha}\left[\frac{1}{x-1}\right] \cdot g(m\|n)\right]$	$t_0 + \frac{t_0}{E_p[n]} \cdot E_p\left[\frac{n}{\overline{g}(n) \cdot \mu_\alpha - 1}\right]$
$P\{T_{\mathcal{M}} \leq TTL\}$	$1 - \frac{1}{E_p[n]} \cdot E_p\left[n \cdot \sum_m E_{m\alpha}\left[\left(\frac{t_0}{TTL}\right)^x\right] \cdot g(m\|n)\right]$	$1 - \frac{1}{E_p[n]} \cdot E_p\left[n \cdot \left(\frac{t_0}{TTL}\right)^{\overline{g}(n)\cdot\mu_\alpha}\right]$



(a) $E[T_{\mathcal{M}}]$ (b) $P\{T_{\mathcal{M}} \leq TTL\}$

Figure 5.2: (a) $E[T_{\mathcal{M}}]$ and (b) $P\{T_{\mathcal{M}} \leq TTL\}$ in scenarios with varying content popularity (α: shape parameter) and $\rho(n) = 0.2 \cdot n$.

with rate λ_{ij})[4]. Then, we create M contents and assign to each of them a popularity value ($N_p^{(\mathcal{M})}$), drawn from the distribution $P_p(n)$. According to the given function $g(m|n)$, we assign the availability values ($N_a^{(\mathcal{M})}$). Finally, for each content $\mathcal{M}$, we randomly choose the $N_p^{(\mathcal{M})}$ nodes that are interested in it and its $N_a^{(\mathcal{M})}$ holders.

Mobility / Popularity patterns: In most of the scenarios we present, we use the gamma distribution for the contact rates (i.e. $f_\lambda(\lambda)$) [107]. Also, content popularity in mobile social networks has been shown to follow a power-law distribution, e.g. [83]. Therefore, we select $P_p(n)$ to follow Discrete (Bounded) Pareto or Zipf distributions, similarly to the majority of related works [34, 85, 99].

In Fig. 5.2 we present the simulation results, along with our theoretical predictions, in scenarios of $N = 10000$ nodes with varying mobility and content popularity patterns. The mean contact rate is $\mu_\lambda = 1$ and content popularity follows a Bounded Pareto distribution with shape parameter (i.e. exponent) α and $n \in [50, 1000]$. The availability function is $\rho(n) = 0.2 \cdot n$

[4]We present here only scenarios where the inter-contact times are exponentially distributed. Similar behavior has been observed in simulations of scenarios with Pareto distributed inter-contact-times.

Table 5.4: Simulation results for scenarios where $g(m|n) \sim Binomial$ with $\overline{g}(n) = 0.2 \cdot n$, and $TTL = 0.05$.

$E[T_{\mathcal{M}}]$ $(x10^3)$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
lower bound	22.3	31.6	52.2	66.4
simulation ($CV_\lambda = 0.5$)	23.9	34.8	57.3	75.0
simulation ($CV_\lambda = 1$)	25.0	36.2	61.9	81.4
$P\{T_{\mathcal{M}} \leq TTL\}$	$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$	$\alpha = 3$
upper bound	0.89	0.81	0.66	0.56
simulation ($CV_\lambda = 2$)	0.87	0.79	0.62	0.52

(i.e. deterministic). An almost perfect match between simulation results (markers) and the theoretical predictions (dashed lines) of Results 11 and 12 can be observed. In Fig. 5.2(a), the lower bound (continuous line) of Theorem 2 is very tight for low mobility (i.e. small CV_λ) and/or content popularity (i.e. small α) heterogeneity. For the delivery probability $P\{T_{\mathcal{M}} \leq TTL\}$ (Fig. 5.2(b)), we present the results for two different values of TTL in scenarios with $CV_\lambda = 2$. Here, the upper bound (continuous line) of Theorem 3 is very close to the simulation results, despite the very heterogeneous mobility.

In Table 5.4 we present results of the above scenarios, where the availability - popularity correlation $g(m|n)$ follows a binomial distribution with mean $\overline{g}(n) = 0.2 \cdot n$. It can be seen that the bounds are tight in most of the scenarios, though (as expected) less tight than in the deterministic $g(m|n)$ case (i.e. $\rho(n)$).

In Fig. 5.3(a) we compare Result 13 with simulations on scenarios conforming to Def. 15: $P_p(n)$ is a *Bounded Pareto* distribution with $\alpha = 2$, and $f_\lambda(\lambda) \sim$ *Pareto*. It can be seen that our theoretical prediction (approximation) achieves good accuracy even in these very heterogeneous mobility scenarios.

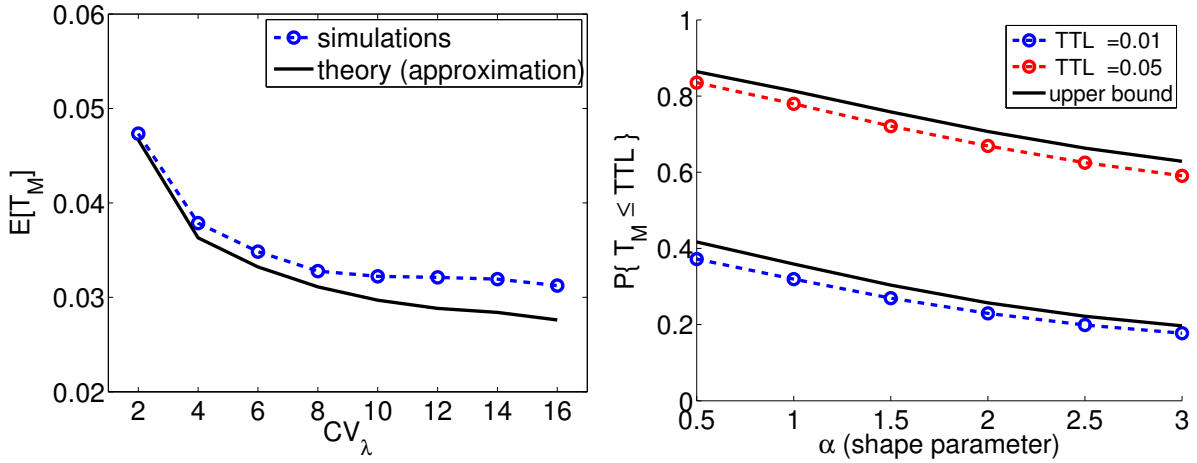
Results for scenarios with mobility-dependent availability (Def. 16) are presented in Fig. 5.3(b). $P_p(n)$ is selected as before and $f_\lambda(\lambda) \sim$ *Gamma* with $\mu_\lambda = 1, CV_\lambda = 0.5$. The allocation of holders is made as in the example in Section 5.3.3.2. The upper bounds of Result 14 are tight in all scenarios, similarly to the case without mobility dependence (Fig. 5.2(b)).

Finally, we need to mention that we have also performed a large number of other scenarios, with similar conclusions.

5.4 Case Study: Mobile Data Offloading

The results of Section 3.3 can be used to predict the performance of a given content allocation policy or content-sharing scheme. In this section, we show how these results could be also used to design / optimize policies. We focus on an application that has recently attracted attention, that of *mobile data offloading* using MSNs [50, 85, 141]. Nevertheless, the same methodology applies for a range of other applications where the number of content/service providers must be chosen.

In a mobile data offloading scenario, the cellular network provider distributes content copies only to some of the nodes interested in this content (holders), in order to reduce the cellular traffic (possibly offering some incentives to the holder nodes). The remaining (interested) nodes must then retrieve the content from the designated holders during direct encounters. A tradeoff



(a) $E[T_{\mathcal{M}}]$ (b) $P\{T_{\mathcal{M}} \leq TTL\}$

Figure 5.3: (a) $E[T_{\mathcal{M}}]$ in scenarios under Def. 15 and (b) $P\{T_{\mathcal{M}} \leq TTL\}$ in scenarios under Def. 16. $\rho(n) = 0.2 \cdot n$.

is involved between the amount of traffic offloaded and the average delay for non-holders. In some cases, an additional *QoS* constraint might exist: if the delay to access a content exceeds a TTL, a requesting node will download it from the infrastructure [50, 85, 141].

Hence, the objective in offloading optimization problems is how the cellular network provider should *choose the set of holders for each content in order to optimize a performance metric*, under a given constraint (e.g. energy or buffer size) and a given popularity distribution $P_p(n)$.

We study cases with and without *QoS* constraints in Sections 5.4.1 and 5.4.2, respectively. For simplicity, we use the expressions of Theorems 2 and 3 as approximations for $E[T_{\mathcal{M}}]$ and $P\{T_{\mathcal{M}} \leq TTL\}$. Since, these expressions imply that (a) the exact mobility patterns are not known (i.e. only μ_λ is needed) and (b) contents with the same popularity are equivalent, our goal is to select the number of holders for each content with a given popularity. In other words, we try to find the optimal allocation function $g(m|n)$.

5.4.1 Case 1: no QoS constraints

When no *QoS* constraints exist, the cellular operator decides the maximum amount of traffic that it wishes to serve directly over the infrastructure. Under this constraint, which can be translated as a constraint on the number of holders for different contents, the objective is to minimize the expected delay $E[T_{\mathcal{M}}]$. The following result, formalizes this optimization problem and provides with the optimal solution for $g(m|n)$.

Result 15. *The minimum expected content access delay, under the constraint of an average number of $c_{\mathcal{M}}$ copies per content, i.e.:*

$$\min\{E[T_{\mathcal{M}}]\} \quad s.t. \quad \sum_{\mathcal{M}} N_a^{(\mathcal{M})} = M \cdot c_{\mathcal{M}} \ , \ N_a^{(\mathcal{M})} \geq 0$$

can be achieved when the allocation function, $g(m|n)$, is deterministic and equal to

$$\rho^*(n) = \frac{c_{\mathcal{M}}}{E_p[\sqrt{n}]} \cdot \sqrt{n}$$

Proof. Using as an approximation for $E[T_{\mathcal{M}}]$ the expression of Theorem 2, we can write

$$E[T_{\mathcal{M}}] = \frac{1}{\mu_\lambda \cdot E_p[n]} \cdot E_p\left[\frac{n}{\overline{g}(n)}\right]$$

Jensen's inequality used in Eq. (5.14), becomes equality when $g(m|n)$ is deterministic. This suggests that among all the functions $g(m|n)$ with the same average value $\overline{g}(n)$, the minimum delay can be achieved in the case: $\rho(n) = \overline{g}(n)$. Thus, the $E[T_{\mathcal{M}}]$ minimization problem becomes equivalent to

$$\min\left\{E_p\left[\frac{n}{\rho(n)}\right]\right\} = \sum_n \frac{n}{\rho(n)} \cdot P_p(n) = \sum_n \frac{n}{\rho_n} \cdot P_p(n) \tag{5.19}$$

where we expressed the expectation as a sum and denoted $\rho_n = \rho(n)$.

Moreover, we can express the content copies constraint as

$$c_{\mathcal{M}} = \frac{\sum_{\mathcal{M}} N_a^{(\mathcal{M})}}{M} = E_p[\rho(n)] = \sum_n \rho_n \cdot P_p(n) \tag{5.20}$$

Using Eq. (5.19) and Eq. (5.20), the optimization problem becomes

$$\min_{\overline{\rho}}\left\{\sum_n \frac{n}{\rho_n} \cdot P_p(n)\right\} \quad s.t. \quad \sum_n \rho_n \cdot P_p(n) = c_{\mathcal{M}} \tag{5.21}$$

where $\overline{\rho}$ denotes the vector with components ρ_n.

The optimization problem of Eq. (5.21) is *convex* and, thus, it can be solved with the method of Lagrange multipliers [4]. Hence, we need to find the values of $\overline{\rho}$ for which it holds that

$$\nabla\left(\sum_n \frac{n}{\rho_n} \cdot P_p(n)\right) + \nabla\lambda_0\left(\sum_n \rho_n \cdot P_p(n) - c_{\mathcal{M}}\right) = 0$$

where λ_0 is the langrangian multiplier. Here, the constraint $\rho_n \geq 0$ needs also to be taken into account. However, it is proved to be an inactive constraint (the solution satisfies it) and thus we omit it at this step for simplicity. Similarly, we assume a large enough network, i.e. always holds $\rho_n \leq N$.

The differentiation over ρ_n gives

$$\rho_n = \frac{1}{\sqrt{\lambda_0}} \cdot \sqrt{n} \tag{5.22}$$

Substituting Eq. (5.22) in the constraint expression $\sum_n \rho_n \cdot P_p(n) = c_{\mathcal{M}}$ (Eq. (5.21)), we can easily get

$$\sqrt{\lambda_0} = \frac{\sum_n \sqrt{n} \cdot P_p(n)}{c_{\mathcal{M}}} = \frac{E_p[\sqrt{n}]}{c_{\mathcal{M}}} \tag{5.23}$$

Then, substituting Eq. (5.23) in Eq. (5.22), gives

$$\rho(n) = \rho_n = \frac{c_{\mathcal{M}}}{E_p[\sqrt{n}]} \cdot \sqrt{n} \tag{5.24}$$

Finally, the values of Eq. (5.24) satisfy the *Karush-Kuhn-Tucker* conditions, which means that the solution of Eq. (5.24) is a global minimum [4]. □

Result 15 is a generic result, since it holds under *any* content popularity pattern. We also note that an allocation policy of $\rho(n) \propto \sqrt{n}$ has also been shown to achieve optimal results in (conventional) peer-to-peer networks [24]. This is an interesting finding, given the inherent differences between the two settings (e.g. node mobility).

Finally, our result is also consistent in scenarios with *mobility dependent holders allocation*. For example, after choosing the number of copies for a content (Result 15), the selection of holders can be made, taking into account mobility utility metrics, e.g. meeting frequency [10] or node centrality [34].

5.4.2 Case 2: QoS constraints

In cases where a maximum delay TTL is required, the objective is to minimize the traffic load served by the infrastructure. The metric used in related work, e.g. [85], is the data offloading ratio, $R_{off.}$, which is defined as the percentage of content requests that are served by nodes. Since requests are served by the infrastructure only after the time TTL elapses, it follows that in our framework: $R_{off.} = P\{T_{\mathcal{M}} \leq TTL\}$.

Hence the optimization problem is equivalent to

$$\max P\{T_{\mathcal{M}} \leq TTL\} \;\; s.t. \;\; \sum_{\mathcal{M}} N_a^{(\mathcal{M})} = M \cdot c_{\mathcal{M}}, \; N_a^{(\mathcal{M})} \geq 0$$

Using the same notation and arguments as in the Section 5.4.1 and the expression of Theorem 3 as an approximation for $P\{T_{\mathcal{M}} \leq TTL\}$, the above optimization problem becomes:

$$\min_{\rho(n)} \left\{ E_p \left[n \cdot e^{-\rho(n)\cdot\mu_\lambda \cdot TTL} \right] \right\} \qquad s.t. \quad E_p[\rho(n)] = c_{\mathcal{M}} \tag{5.25}$$

with $\rho(n) \geq 0$, or, equivalently:

$$\min_{\overline{\rho}} \quad \left\{ \sum_n n \cdot e^{-\rho_n \cdot \mu_\lambda \cdot TTL} \cdot P_p(n) \right\} \qquad s.t. \quad \sum_n \rho_n \cdot P_p(n) = c_{\mathcal{M}} \;, \quad \rho_n \geq 0 \tag{5.26}$$

The optimization problem of Eq. (5.26) is convex. Although a closed form solution, as in Result 15, cannot be derived in the same way[5], it can be solved numerically, using well known methods.

5.4.3 Performance Evaluation

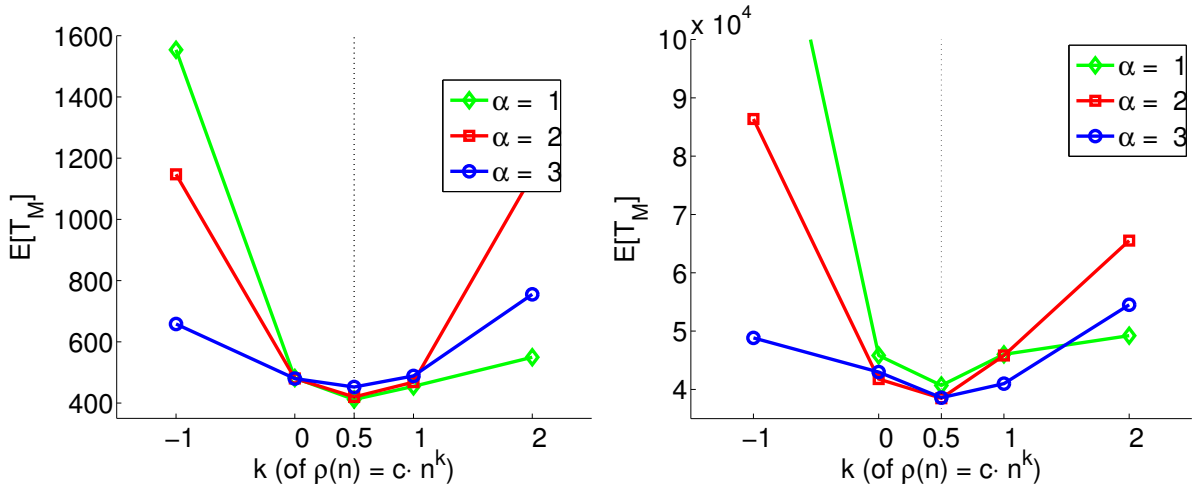
To investigate whether the policies suggested as optimal by our theory indeed perform better, we conducted simulations on various synthetic scenarios and on traces of real networks, where node mobility patterns usually involve much more complex characteristics than our model.

The results in the majority of scenarios considered have been encouragingly consistent with our theoretical predictions. Hence, we only present here a small, representative sample. Specifically, we consider the following traces coming from state-of-the-art mobility models or collected in experiments.

TVCM mobility model [57]: Scenario with 100 nodes divided in 4 communities of unequal size. Nodes move mainly inside their community and leave it for a few short periods.

SLAW mobility model [77]: Network with 200 nodes moving in a square area of $2000m$ (the

[5]The difference here is that the constraint $\rho_n \geq 0$ is active.



(a) TVCM (b) Cabspotting

Figure 5.4: Content access delay $E[T_{\mathcal{M}}]$ of different allocation policies $\rho(n) = c_k \cdot n^k$, where $c_k = \frac{c_{\mathcal{M}}}{E_p[n^k]}$.

other parameters are set as in the source code provided in [77]).
Cabspotting trace [118]: GPS coordinates from 536 taxi cabs collected over 30 days in San Francisco. A range of $100m$ is assumed.
Infocom trace [58]: Bluetooth sightings of 98 mobile and static nodes (iMotes) collected in an experiment during Infocom 2006.

5.4.3.1 Case 1: no QoS constraints

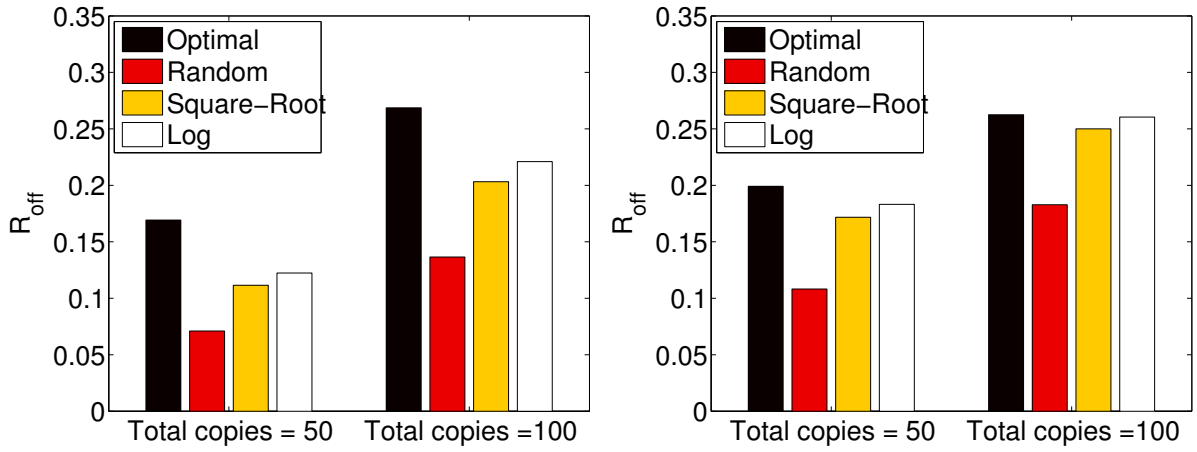
In each scenario, we compare different allocation functions $\rho(n) = c_k \cdot n^k$, where $c_k = \frac{c_{\mathcal{M}}}{E_p[n^k]}$ is a normalization factor such that the constraint $E_p[\rho(n)] = c_{\mathcal{M}}$ is satisfied.

In Fig. 5.4 we present simulation results in scenarios for the *TVCM* (Fig. 5.4(a)) and Cabspotting (Fig. 5.4(b)) traces. Content popularity ($P_p(n)$) follows a *Zipf* distribution with $n \leq 30$ and exponent $\alpha = \{1, 2, 3\}$. The availability constraint is set to $c_{\mathcal{M}} = 10$. It can be seen that the optimal delay $E[T_{\mathcal{M}}]$ is achieved for $k = 0.5$, as Result 15 predicts (despite the fact that we used the expression of the lower bound as an approximation for the expected delay $E[T_{\mathcal{M}}]$).

5.4.3.2 Case 2: QoS constraints

To evaluate the performance of the allocation function $\rho(n)$ that follows after solving Eq. (5.26) (i.e. *optimal* allocation), we compare the *offloading ratio* R_{off} it achieves with the offloading ratios of the following policies:
Random: We randomly select a content and give a copy of it to a node. We repeat $M \cdot c_{\mathcal{M}}$ times.
Square Root: We select $\rho(n) \propto \sqrt{n}$ (i.e. the allocation that achieves the minimum expected delay $E[T_M]$).
Log: We select $\rho(n) \propto \log n$.

Random policy has been used in related work as a baseline [85] and *square root* policy is the optimal policy when the metric of interest is the content access delay (Section 5.4.1). Finally,



(a) SLAW, $TTL = 530$ (b) Infocom, $TTL = 10000$

Figure 5.5: Offloading Ratio $R_{off.}$ of different allocation policies $\rho(n)$.

we observed that the *optimal* policy (Eq. (5.26)), in the scenarios considered, allocated copies only to the $10\% - 20\%$ highest popularity contents. The *log* policy allocates in a similar manner the copies (e.g. no copies to contents with low popularity).

Simulation results on the *SLAW* and *Infocom* scenarios are presented in Fig. 5.5(a) and 5.5(b), respectively. The parameters in these scenarios are: $M = 50$ messages, $P_p \sim Zipf$ with $n \in [1, 30]$ and $\alpha = 1$, total copies $M \cdot c_{\mathcal{M}} = \{50, 100\}$. As it can be seen our *optimal* policy (leftmost bar) achieves the highest offloading ratio $R_{off.}$. The random policy is clearly inferior than the others. Between *square root* and *log* policies, it is the latter that achieves better performance. These results indicate that, to maximize $R_{off.}$, it is better to allocate the available resources only for popular contents, and serve the non-popular exclusively through the infrastructure.

5.5 Related Work

Content-centric applications were introduced in mobile social networking under the *publish - subscribe* paradigm [10, 28, 79, 142], for which several data dissemination techniques have been proposed. In [142], authors propose a mechanism that identifies social communities and the nodes-"hubs", and builds an overlay network between them in order to efficiently disseminate data. SocialCast [28] based on information about nodes interests, social relationships and movement predictions, selects the set of holders. Similarly to the above approaches, ContentPlace [10] uses both community detection and nodes social relationships information, to improve the performance of the content distribution.

Under a different setting, [34, 99] study content sharing mechanisms with limited resources (e.g. buffer sizes, number of holders). In [34], authors analytically investigate the data dissemination cost-effectiveness tradeoffs, and propose techniques based on contact patterns (i.e. λ_{ij}) and nodes interests. Similarly, CEDO [99] aims at maximizing the total content delivery rate: by maintaining a utility per content, nodes make appropriate drop and scheduling decisions.

Recently, further novel content-centric application have been proposed, like location-based

applications [105, 123] and mobile data offloading [50, 85, 141]. The latter category, due to the rapid increase of mobile data demand, has attracted a lot of attention. In the setting of [141], content copies are initially distributed (through the infrastructure) to a subset of mobile nodes, which then start propagating the contents epidemically. Differently, in [50] the authors consider a limited number of holders, and study how to select the best holders-target-set for each message. In [85], the same problem is considered, and (centralized) optimization algorithms are proposed that take into account more information about the network: namely, size and lifetimes of different contents, and interests, privacy policies and buffer sizes of each node.

In the majority of previous studies, although node interests and content popularity are taken into account, the focus has been on the algorithms and the applications themselves. We believe that our study complements existing work, by providing a common analytical framework for a number of these approaches that can be used both for predicting the performance of proposed schemes, as well as proposing improved ones.

5.6 Conclusion

The increasing number of mobile devices and traffic demand, renders content-centric applications through opportunistic communication very promising. Hence, motivated by the lack of a common analytical framework, we modeled and analyzed the effects of content popularity / availability patterns in the performance of content-centric mechanisms. In the future, we believe that the effect of further dimensions of content-centric traffic should be investigated, like *traffic locality*, independently from or jointly with the popularity dimension.

Based on this first analysis of content-centric traffic patterns, and the insights obtained from our results, in the following chapter, we design an MSN-related content-centric application for offloading mobile data traffic from cellular networks, analyze its performance and study a number of optimization approaches.

5.7 Appendix: Supplementary Theoretical Results and Proofs

5.7.1 Proof of Result 13

Proof. To calculate the average performance, we need to modify the previous analysis as following: Consider a content $\mathcal{M}$ of initial popularity $N_p^{(\mathcal{M})}(0) = n$ and availability $N_a^{(\mathcal{M})}(0) = m$, i.e. initially n nodes are looking for the content and m nodes hold the content. When the first requester access the content, the number of holders will increase to $m + 1$ and the remaining requesters will be $n - 1$. Building a Markov Chain as in Fig 5.1, where each state denotes the number of holders, it can be shown for the expected delay of moving from state $m + k$ to state $m + k + 1$, $k \in [0, 1]$, that it holds

$$E[T_{k,k+1}] \approx \frac{1}{(m+k) \cdot (n-k) \cdot \mu_\lambda} \tag{5.27}$$

where $m+k$ are the nodes holding the content, $n-k$ the remaining requesters and μ_λ the mean contact rate.

From the above analysis, it follows straightforward that the expected time till the first

requester to access the message is $E[T^1] = E[T_{0,1}]$ and till the ℓ^{th} requester to access it is

$$E[T^\ell] = \sum_{k=0}^{\ell-1} E[T_{k,k+1}] \tag{5.28}$$

Let us now define the sum of delays $E[T^\ell]$ (i.e. delivery delays for each requester) for a message $\mathcal{M}$ with initial availability $N_a^{(\mathcal{M})}(0) = m$ and initial popularity $N_p^{(\mathcal{M})}(0) = n$, as:

$$S(T_\mathcal{M}|m,n) = \sum_{\ell=1}^{n} E[T^\ell|m,n] \tag{5.29}$$

From Eq. (5.27) and Eq. (5.28), we can write for $S(T_\mathcal{M}|m,n)$:

$$\begin{aligned} S(T_\mathcal{M}|m,n) \approx \sum_{\ell=1}^{n}\sum_{k=0}^{\ell-1} \frac{1}{(m+k)\cdot(n-k)\cdot\mu_\lambda} &= \sum_{k=0}^{n-1}(n-k)\cdot\frac{1}{(m+k)\cdot(n-k)\cdot\mu_\lambda} \\ &= \frac{1}{\mu_\lambda}\cdot\sum_{k=0}^{n-1}\frac{1}{m+k} = \frac{1}{\mu_\lambda}\cdot\sum_{k=m}^{m+n-1}\frac{1}{k} \end{aligned} \tag{5.30}$$

and using the approximation of the harmonic sum[6], we get

$$S(T_\mathcal{M}|m,n) \approx \frac{1}{\mu_\lambda}\cdot\ln\left(1+\frac{n}{m-1}\right) \approx \frac{1}{\mu_\lambda}\cdot\ln\left(1+\frac{n}{m}\right) \tag{5.31}$$

Averaging over all the content in the network, we can write for the expected content access delay:

$$E[T_\mathcal{M}] = \frac{\sum_\mathcal{M} S(T_\mathcal{M})}{\sum_\mathcal{M} N_p^{(\mathcal{M})}} \tag{5.32}$$

or, since (i) (by definition) there are $M\cdot P_p(n)$ contents in the network, and (ii) we do not differentiate between contents with the same popularity/availability:

$$\begin{aligned} E[T_\mathcal{M}] = \frac{\sum_n S(T_\mathcal{M}|n)\cdot(M\cdot P_p(n))}{\sum_\mathcal{M} n\cdot(M\cdot P_p(n))} &= \frac{\sum_n S(T_\mathcal{M}|n)\cdot P_p(n)}{E_p[n]} \\ = \frac{\sum_n S(T_\mathcal{M}|n,m)\cdot g(m|n)\cdot P_p(n)}{E_p[n]} &\approx \frac{\sum_n \frac{1}{\mu_\lambda}\cdot\ln\left(1+\frac{n}{m}\right)\cdot g(m|n)\cdot P_p(n)}{E_p[n]} \end{aligned} \tag{5.33}$$

where in the last line we substituted from Eq. (5.31).

We can further use Jensen's inequality (since the function $h(x) = \ln\left(1+\frac{n}{x}\right)$ is convex) or the respective approximation, and finally write:

$$E[T_\mathcal{M}] \approx \frac{1}{\mu_\lambda\cdot E_p[n]}\cdot E_p\left[\ln\left(1+\frac{n}{\overline{g}(n)}\right)\right] \tag{5.34}$$

which proves the result. □

[6] $\sum_{k=1}^{N} \approx \ln(N) + \gamma + O\left(\frac{1}{N}\right)$, where γ is the *Euler-Mascheroni* constant.

5.7.2 Proof of Result 14 and Example

Proof. Def. 16 says that *who* holds a content and *who* is interested in it is not independent of their mobility patters. The contact rates between the requester of a content and the holders of it, are not distributed with the contact rates distribution $f_\lambda(\lambda)$, since the requesters-holders contact rates are mobility dependent. It can be shown (see Chapter 4; Result 1) that the requesters-holders contact rates are distributed as

$$f_\pi(\lambda) = \frac{1}{E_\lambda[\pi(\lambda)]} \cdot \pi(\lambda) \cdot f_\lambda(\lambda) \tag{5.35}$$

Hence, Eq. (5.7) and Eq. (5.8) need to be modified as:

$$X_\mathcal{M} \sim f_{m\pi}(x) = (f_\pi * f_\pi \cdots * f_\pi)_m \tag{5.36}$$

and

$$E[X_\mathcal{M}|N_a^{(\mathcal{M})} = m] = E_{m\pi}[x] = m \cdot \frac{E_\lambda[\lambda \cdot \pi(\lambda)]}{E_\lambda[\pi(\lambda)]} = m \cdot \mu_\lambda^{(\pi)} \tag{5.37}$$

Then, it can be easily seen that following the same analysis, we get the same expressions as in Theorems 2 and 3 and Result 13 where, now, the mean contact rate μ_λ is replaced by the *mean mobility dependent requesters-holders contact rate* $\mu_\lambda^{(\pi)}$. □

Example Scenario: For each content $\mathcal{M}$, its holders are selected taking into account their contact rates with the requesters with the following mechanism: Each node i candidate to be a holder is assigned a weight $w_i = \prod_{j \in \mathcal{C}_p^{(\mathcal{M})}} \lambda_{ij}$. Then, each of them is selected to be one of the $N_a^{(\mathcal{M})}$ holders with probability $p_i = \frac{w_i}{\sum_i w_i}$. Now, for the node pair $\{i, j\}$ ($i \in \mathcal{C}_a^{(\mathcal{M})}$, $j \in \mathcal{C}_p^{(\mathcal{M})}$) it holds that

$$\pi_{ij} = \frac{w_i}{\sum_i w_i} = \frac{\prod_{k \in \mathcal{C}_p^{(\mathcal{M})}} \lambda_{ik}}{\sum_i \prod_{k \in \mathcal{C}_p^{(\mathcal{M})}} \lambda_{ik}} = \frac{\lambda_{ij} \cdot \prod_{k \in \mathcal{C}_p^{(\mathcal{M})} \setminus \{j\}} \lambda_{ik}}{\sum_i \prod_{k \in \mathcal{C}_p^{(\mathcal{M})}} \lambda_{ik}} \tag{5.38}$$

for which, when the node popularity $N_p^{(\mathcal{M})} = |\mathcal{C}_p^{(\mathcal{M})}|$ is large enough, we can write

$$\pi_{ij} \approx \frac{\lambda_{ij} \cdot c_1}{c_2} \tag{5.39}$$

where c_1, c_2 take approximately the same value $\forall i, j$, i.e. $\pi(\lambda) = c \cdot \lambda$, $c = \frac{c_1}{c_2}$. Substituting $\pi(\lambda)$ in Result 14, gives

$$\mu_\lambda^{(\pi)} = \frac{E_\lambda[\lambda \cdot \pi(\lambda)]}{E_\lambda[\pi(\lambda)]} = \frac{E_\lambda[\lambda^2]}{E_\lambda[\lambda]} = \mu_\lambda \cdot (1 + CV_\lambda^2) \tag{5.40}$$

5.7.3 Minimum of Pareto distributed random variables

For the random variable $T_\mathcal{M} = \min_{i \in \mathcal{C}_a^{(\mathcal{M})}} \{T_{ij}\}$, where each T_{ij} is a random variable distributed with a Pareto distribution with scale parameter t_0 and shape parameter α_{ij}, it holds that:

$$P\{T_\mathcal{M} > t\} = \prod_{i \in \mathcal{C}_a^{(\mathcal{M})}} P\{T_{ij} > t\} = \prod_{i \in \mathcal{C}_a^{(\mathcal{M})}} \left(\frac{t_0}{t}\right)^{\alpha_{ij}} = \left(\frac{t_0}{t}\right)^{\sum_{i \in \mathcal{C}_a^{(\mathcal{M})}} \alpha_{ij}} \tag{5.41}$$

which means that $T_\mathcal{M}$ follows a Pareto distribution with scale and shape parameters t_0 and $A_\mathcal{M} = \sum_{i \in \mathcal{C}_a^{(\mathcal{M})}} \alpha_{ij}$, respectively

5.7.4 Proofs for the performance metrics expressions of the Pareto case

5.7.4.1 Content Access Delay

The expectation of a Pareto random variable ($pareto(t_0, \alpha_{ij})$) is $\frac{t_0\alpha}{\alpha-1}$. Hence, in the derivation of Eq. (5.9), one only needs to change the integral in the last equality as:

$$E[T_{\mathcal{M}}] = \sum_m \int \frac{x \cdot t_0}{x-1} \cdot f_{m\alpha}(x)dx \cdot P_a^{req.}(m) \tag{5.42}$$

Substituting $P_a^{req.}(m)$ from Lemma 8 and proceeding as in the exponential case, we subsequently get:

$$\begin{aligned} E[T_{\mathcal{M}}] &= \sum_m \int \frac{x \cdot t_0}{x-1} \cdot f_{m\alpha}(x)dx \cdot \frac{E_p[n \cdot g(m|n)]}{E_p[n]} = \frac{t_0}{E_p[n]} \cdot E_p\left[n \cdot \sum_m E_{m\alpha}\left[\frac{x}{x-1}\right] \cdot g(m|n)]\right] \\ &= \frac{t_0}{E_p[n]} \cdot E_p\left[n \cdot \sum_m \left(1 + E_{m\alpha}\left[\frac{1}{x-1}\right]\right) \cdot g(m|n)]\right] \\ &= \frac{t_0}{E_p[n]} \cdot E_p\left[n + n \cdot \sum_m E_{m\alpha}\left[\frac{1}{x-1}\right] \cdot g(m|n)]\right] \\ &= t_0 + \frac{t_0}{E_p[n]} \cdot E_p\left[n \cdot \sum_m E_{m\alpha}\left[\frac{1}{x-1}\right] \cdot g(m|n)\right] \end{aligned} \tag{5.43}$$

which is the exact expression for $E[T_{\mathcal{M}}]$ in Table 5.3.

Applying *Jensen's inequality* for the convex function $h(x) = \frac{1}{x-1}$, gives:

$$E_{m\alpha}\left[\frac{1}{x-1}\right] \geq \frac{1}{m \cdot \mu_\alpha - 1} \tag{5.44}$$

and, thus:

$$\begin{aligned} E[T_{\mathcal{M}}] \geq t_0 + \frac{t_0}{E_p[n]} \cdot E_p\left[n \cdot \sum_m \frac{1}{m \cdot \mu_\alpha - 1} \cdot g(m|n)\right] &= t_0 + \frac{t_0}{E_p[n]} \cdot E_p\left[n \cdot E_g\left[\frac{1}{m \cdot \mu_\alpha - 1}\right]\right] \\ &\geq t_0 + \frac{t_0}{E_p[n]} \cdot E_p\left[n \cdot \frac{1}{\overline{g}(n) \cdot \mu_\alpha - 1}\right] \end{aligned} \tag{5.45}$$

where for the last line we applied *Jensen's inequality* for the expectation $E_g\left[\frac{1}{m \cdot \mu_\alpha - 1}\right]$.

5.7.4.2 Content Access Probability

In the Pareto case, the integral in Eq. (5.15) changes as: $\int \left(\frac{t_0}{TTL}\right)^x \cdot f_{m\alpha}(x)dx$, for $TTL \geq t_0$, because for a Pareto random variable $x \sim pareto(t_0, \alpha)$ it holds that $P\{x \leq TTL\} = 1 - \left(\frac{t_0}{TTL}\right)^\alpha$. Following the same methodology as before and observing that the function $h(x) = \left(\frac{t_0}{TTL}\right)^\alpha$ is convex, the expressions of Table 5.3 follow similarly.

5.7.5 Proof of Expressions in Table 5.2

Meeting rates are distributed with a Gamma distribution $f_\lambda(x) = \Gamma(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} \cdot e^{-\beta \cdot x}$, where $\Gamma(\alpha)$ is the *gamma function* and the parameters α, β relate to the expectation μ_λ and coefficient of variation CV_λ of the meeting rates, as following:

$$\alpha = \frac{1}{CV_\lambda^2} \quad \text{and} \quad \beta = \frac{1}{\mu_\lambda \cdot CV_\lambda^2} \tag{5.46}$$

The sum of m random variables $\lambda_{ij} \sim f_\lambda(x) = \Gamma(x;\alpha,\beta)$ is a random variable and is distributed with a gamma distribution with parameters $m \cdot \alpha$ and β, i.e.

$$f_{m\lambda} = \Gamma(x; m \cdot \alpha, \beta) = \frac{\beta^{m\cdot\alpha}}{\Gamma(m \cdot \alpha)} \cdot x^{m\cdot\alpha-1} \cdot e^{-\beta \cdot x} \tag{5.47}$$

Moreover, when

$$P_p(n) \sim Pareto(x; n_0, \alpha = 2) = \frac{2 \cdot n_0^2}{x^3}, \quad for\ x \geq n_0 \tag{5.48}$$

the mean value of the popularity is

$$E_p[n] = \frac{2 \cdot n_0}{2-1} = 2 \cdot n_0 \tag{5.49}$$

5.7.5.1 Delivery Delay $E[T_\mathcal{M}]$

To calculate the delivery delay, given by Result 11, we need first to calculate the quantity $E_{m\lambda}\left[\frac{1}{x}\right]$. From Eq. (5.47) we get

$$\begin{aligned} E_{m\lambda}\left[\frac{1}{x}\right] &= \int_0^\infty \frac{1}{x} \cdot \frac{\beta^{m\cdot\alpha}}{\Gamma(m\cdot\alpha)} \cdot x^{m\cdot\alpha-1} \cdot e^{-\beta\cdot x} \cdot dx = \int_0^\infty \frac{\beta^{m\cdot\alpha}}{\Gamma(m\cdot\alpha)} \cdot x^{(m\cdot\alpha-1)-1} \cdot e^{-\beta\cdot x} \cdot dx \\ &= \frac{\beta\cdot\Gamma(m\cdot\alpha-1)}{\Gamma(m\cdot\alpha)} \int_0^\infty \frac{\beta^{m\cdot\alpha-1}}{\Gamma(m\cdot\alpha-1)} \cdot x^{(m\cdot\alpha-1)-1} \cdot e^{-\beta\cdot x} \cdot dx \overset{*}{=} \frac{\beta\cdot\Gamma(m\cdot\alpha-1)}{\Gamma(m\cdot\alpha)} \\ &\overset{**}{=} \frac{\beta}{m\cdot\alpha-1} = \frac{1}{m\cdot\frac{\beta}{\alpha}-\beta} \overset{***}{=} \frac{1}{\mu_\lambda}\frac{1}{m-CV_\lambda^2} \end{aligned} \tag{5.50}$$

because it holds that

$$\begin{aligned} &(*) \quad \int_0^\infty \frac{\beta^{m\cdot\alpha-1}}{\Gamma(m\cdot\alpha-1)} \cdot x^{(m\cdot\alpha-1)-1} \cdot e^{-\beta\cdot x} \cdot dx = \int_0^\infty \Gamma(x; m\cdot\alpha-1, \beta) \cdot dx = 1 \\ &(**) \quad \Gamma(\zeta+1) = \Gamma(\zeta) \end{aligned}$$

and for (***) we used Eq. (5.46).

Substituting Eq. (5.50) and $g(m|n) \equiv \rho(n) = c \cdot n$ in the expression of Result 11, we get

$$E[T_\mathcal{M}] = \frac{1}{E_p[n]} \cdot E_p\left[n \cdot \frac{1}{\mu_\lambda} \cdot \frac{1}{c\cdot n - CV_\lambda^2}\right] \tag{5.51}$$

To calculate the expectation $E_p\left[n \cdot \frac{1}{\mu_\lambda}\frac{1}{c\cdot n - CV_\lambda^2}\right]$ we use Eq. (5.48)

$$E_p\left[n \cdot \frac{1}{\mu_\lambda} \cdot \frac{1}{c\cdot n - CV_\lambda^2}\right] = \int_{n_0}^{\infty} x \cdot \frac{1}{\mu_\lambda} \cdot \frac{1}{c\cdot x - CV_\lambda^2} \cdot \frac{2\cdot n_0^2}{x^3} \cdot dx = \frac{2\cdot n_0^2}{\mu_\lambda}\int_{n_0}^{\infty} \frac{1}{x^2\cdot(c\cdot x - CV_\lambda^2)} \cdot dx$$
$$= \frac{2\cdot n_0^2}{\mu_\lambda \cdot CV_\lambda^4}\left[\frac{c\cdot x \cdot \ln\left(c - \frac{CV_\lambda^2}{x}\right) + CV_\lambda^2}{x}\right]_{n_0}^{\infty} = \frac{2\cdot n_0^2}{\mu_\lambda \cdot CV_\lambda^4}\left[c\cdot\ln(c) - c\cdot\ln\left(c - \frac{CV_\lambda^2}{n_0}\right) - \frac{CV_\lambda^2}{n_0}\right]$$
$$= \frac{2\cdot n_0^2}{\mu_\lambda \cdot CV_\lambda^4}\left[c\cdot\ln\left(\frac{c}{c - \frac{CV_\lambda^2}{n_0}}\right) - \frac{CV_\lambda^2}{n_0}\right] = \frac{2\cdot n_0}{\mu_\lambda \cdot CV_\lambda^2}\left[\frac{c\cdot n_0}{CV_\lambda^2}\cdot\ln\left(\frac{1}{1 - \frac{CV_\lambda^2}{c\cdot n_0}}\right) - 1\right] \quad (5.52)$$

Finally, substituting Eq. (5.49) and Eq. (5.52) in Eq. (5.51), we get the expression of Table 5.2.

5.7.5.2 Delivery Probability $P\{T_{\mathcal{M}} \leq TTL\}$

We follow a similar procedure as above.

To calculate the delivery probability, given by Result 12, we need first to calculate the quantity $E_{m\lambda}\left[e^{-x\cdot TTL}\right]$. From Eq. (5.47) we get

$$E_{m\lambda}\left[e^{-x\cdot TTL}\right] = \int_0^{\infty} e^{-x\cdot TTL}\cdot\frac{\beta^{m\cdot\alpha}}{\Gamma(m\cdot\alpha)}\cdot x^{m\cdot\alpha-1}\cdot e^{-\beta\cdot x}\cdot dx$$
$$= \int_0^{\infty}\frac{\beta^{m\cdot\alpha}}{\Gamma(m\cdot\alpha)}\cdot x^{m\cdot\alpha-1}\cdot e^{-(\beta+TTL)\cdot x}\cdot dx$$
$$= \frac{\beta^{m\cdot\alpha}}{(\beta+TTL)^{m\cdot\alpha}} = \frac{1}{(1+\frac{TTL}{\beta})^{m\cdot\alpha}} = \left(1+\mu_\lambda\cdot CV_\lambda^2\cdot TTL\right)^{-\frac{m}{CV_\lambda^2}} \quad (5.53)$$

Substituting Eq. (5.53) and $g(m|n) \equiv \rho(n) = c\cdot\ln(n)$ in the expression of Result 12, we get

$$E[T_{\mathcal{M}}] = 1 - \frac{E_p\left[x\cdot\left(1+\mu_\lambda\cdot CV_\lambda^2\cdot TTL\right)^{-\frac{c\cdot\ln(x)}{CV_\lambda^2}}\right]}{E_p[n]} \quad (5.54)$$

To calculate the expectation

$$E_p\left[x\cdot\left(1+\mu_\lambda\cdot CV_\lambda^2\cdot TTL\right)^{-\frac{c\cdot\ln(x)}{CV_\lambda^2}}\right] = E_p\left[x\cdot\gamma_1^{ln(x)}\right] \quad (5.55)$$

where we denoted

$$\gamma_1 = \left(1+\mu_\lambda\cdot CV_\lambda^2\cdot TTL\right)^{-\frac{c}{CV_\lambda^2}} \quad (5.56)$$

we use Eq. (5.48):

$$E_p\left[x\cdot\gamma_1^{ln(x)}\right] = \int_{n_0}^{\infty} x\cdot\gamma_1^{ln(x)}\cdot\frac{2\cdot n_0^2}{x^3}\cdot dx = 2\cdot n_0^2\int_{n_0}^{\infty}\frac{\gamma_1^{ln(x)}}{x^2}\cdot dx = \frac{2\cdot n_0^2}{1-\ln(\gamma_1)}\cdot\left[\frac{\gamma_1^{ln(x)}}{x}\right]_{\infty}^{n_0} \quad (5.57)$$

and for $\gamma_1 < e$, it gives

$$E_p\left[x \cdot \gamma_1^{ln(x)}\right] = \frac{2 \cdot n_0^2}{1 - \ln(\gamma_1)} \cdot \frac{\gamma_1^{ln(n_0)}}{n_0} = \frac{2 \cdot n_0}{1 + \ln(\gamma)} \cdot \frac{1}{\gamma^{ln(n_0)}} \tag{5.58}$$

where $\gamma = \frac{1}{\gamma_1}$.

Finally, substituting Eq. (5.49) and Eq. (5.58) in Eq. (5.54), we get the expression of Table 5.2.

Chapter 6

Offloading on the Edge: Analysis and Optimization of Local Data Storage and Offloading in HetNets

6.1 Introduction

The growth in the number of mobile devices and connection speeds has led to a high volume of mobile data traffic. Cellular networks are currently overloaded and, despite a lot of planned improvements on the physical layer technologies, they are not expected to be able to keep up with the rapidly increasing user data demand [23]. Radically reducing the communication distance by deploying, and *offloading* traffic to, many "small cells" (e.g. femto, pico, or even WiFi) is seen as the only viable solution [3,21,78]. Nevertheless, this requires a large investment in upgrading the backhaul network, increasingly based on wireless links, which will often be the new performance bottleneck [126]. Caching popular content at the "edge", i.e. on storage devices installed at small cell base stations could alleviate backhaul congestion [120, 126], and is supported by a number of real data studies suggesting a high amount of demand overlap between user requests [32, 38, 87].

Reducing the communication distance is taken yet a step further with the newly proposed paradigm of device-to-device (D2D) communication [5, 65]. A device can store a (popular) content after consuming it, and give it directly to other neighboring devices also interested in it, offloading these requests from the main network. The connection between the two devices could be in-band (cellular frequencies) or out of band (e.g. Bluetooth, WiFi Direct). While D2D-based offloading normally assumes a content request will either be served *immediately* from a device currently in range or the cellular network, some recent works have suggested the use of *opportunistic offloading* through D2D: a device requesting some content might wait for some amount of time until it *encounters* another device sharing the content [85,124,139], and go back to the main network if not found before some set deadline.

Hence, more data could be offloaded from the main network through such D2D communication, perhaps at the expense of increased delay for some requests. Such increased delays could sometimes be acceptable (e.g. asynchronous requests, longer start-up or buffering delays easily amortized when considering large content). Yet, in many cases, the operator will need to provide appropriate incentives to these users, either in the form of instantaneous price reductions [48] or

low(er) priced plans. What is more, operators will probably need to also provide incentives to the devices storing the content and acting as local *relays* on their behalf, as this raises important battery consumption, storage, as well as privacy and security issues.

The provision of these incentives constitutes another important form of cost for the operator, together with the costs of directly serving the content from the main (mostly macro-cell based) network, and that of installing, maintaining, and supporting with ample backhaul capacity, new small cells with large enough caches. It thus becomes increasingly important for an operator of such a future Heterogeneous Network (HetNet) with caching and D2D capabilities to be able to answer questions like: *"How much content can be offloaded by a given setup as a function of content demand patterns?"*, *"Is it worth investing in additional cell densification, or would it be more cost-efficient to provide incentives for D2D opportunistic offloading?"*.

To this end, in this chapter we propose an analytical model for studying the problem of "offloading on the edge" in a HetNet. Although capturing all the fine details of possible setups and technologies would be a rather daunting task, we assume two main mechanisms being employed in the considered network, namely (i) caching on small cells and mobile devices, collectively referred to as "edge nodes", and (ii) offloading requests through local, short range communications (e.g. D2D or low power communication to local femto or pico base stations). We first describe the "offloading on the edge" mechanism and propose a generic model that allows us to analytically study it (Section 6.2). We proceed by deriving useful results for the performance of content delivery through this mechanism and the incurred costs, as a function of key system parameters (Section 6.3). Then, we study the total offloading cost and provide insights for content placement and dissemination strategies that minimize this cost (Section 6.4). Finally, we validate our results through realistic simulations (Section 6.5) and discuss related work (Section 6.6).

Summarizing, the main contributions of our work are:

- To our best knowledge, this is the first work jointly and analytically studying offloading through small cells, opportunistic D2D, and caching at both.
- We provide closed-form analytical approximations applicable to a number of performance metrics and network setups.
- We provide initial insights into the various design tradeoffs involved, as well as the efficient allocation of storage space among different contents.

6.2 Offloading on the Edge

6.2.1 Network Setup

We consider a Heterogeneous Cellular Network (HetNet) [3], composed of 3 sets of nodes:

Macro-cell Base Stations ($\mathcal{BS}$): They provide full coverage to subscribed mobile nodes (MNs), but we assume their radio resources are congested.

Small Cells ($\mathcal{SC}$): These are shorter range, low power base stations (e.g. femto and pico-cells, or even WiFi access points) dispersed in the area of coverage. They provide ample capacity to the few MNs within range, and their communication cost to/from a MN is smaller [66]. Hence, they can be used to offload some traffic from BSs. However, the backhaul connection for these cells will often be wireless (either to a BS or to an aggregation point) and underprovisioned [126].

This makes a backhaul transmission to a small cell costly. To this end, each small cell is equipped with some storage capacity, as in [120, 126], where (popular) content could be cached to avoid duplicate backhaul accesses.

Mobile Nodes ($\mathcal{MN}$): These include smartphones, tablets, netbooks, etc. MNs can communicate with BSs, SCs (if in range), and even other MNs directly, if D2D communication is allowed. D2D communication potentially offers higher rates at lower interference levels [5]. Yet, appropriate incentives from the operator might be needed. Without loss of generality, we assume out-of-band communication (e.g. WiFi Direct or Bluetooth) for D2D. We also assume that each MN also has some storage capacity (normally less than that of a small cell) for caching (popular) content.

The number of nodes in each set is

$$N_{BS} = |\mathcal{BS}| \quad , \quad N_{SC} = |\mathcal{SC}| \quad , \quad N_{MN} = |\mathcal{MN}|$$

where $|\cdot|$ denotes the cardinality of a set.

6.2.2 Offloading Mechanism

Content Requests. We assume that each MN is interested in different contents over time (e.g. videos, web pages, software updates, etc.), and that the same content may be of interest to multiple MNs. This interest overlap is supported by recent studies (e.g. [32, 38, 87], to name a few), where the popularity distribution of contents is shown to be highly skewed. In the remainder, we will be assuming that the number of nodes interested in a content, the content popularity, is known in advance or can be estimated. For a number of applications, like *push services* [139], this information can be known in advance by the cellular network. Users are subscribed to a push service they are interested in (e.g. news, series episodes, trending videos, etc.), and when a content (of this service) is created or published, the content provider starts distributing (*pushing*) it to them[1]. Similarly, users might subscribe to certain categories of contents, such as personalized Internet radio stations like Pandora and Jango[2]. The content of these pseudo-random streams of songs can be decided in advance, and thus the popularity of songs belonging to different streams can be estimated.

Content Delivery. An operator can deliver a content to an interested MNs in one of the following ways: (i) *Direct transmission* from a BS; (ii) *Offloading through SCs and/or MNs*, where the operator transmits the content to some SCs over the backhaul and stores it there, or instructs some MNs to store a content for some time (e.g. keeping in their cache a content they consumed). Then, the operator can ask an interested MN within range of a SC or MN caching that content to retrieve it directly.

Moreover, an operator can ask an MN interested in a content θ, but not *currently* within range of an SC or MN with content θ in its cache, to wait for an amount of time, let TTL, until it *moves* within range of such an SC or MN. If this time expires, then the operator is obliged to deliver the content directly through the closest macro BS. While this *delay-tolerant* approach is in contrast to the usual ones considered for small cell and D2D based offloading [65, 120, 126], it is likely that the small cell and (D2D enabled) mobile node density will not always be enough to

[1] We assume that the content provider may be the cellular network operator itself or in cooperation with it (like the Akamai and Swisscom example [1]).

[2] `www.pandora.com` , `www.jango.com`

offload enough traffic. Hence, it is a valuable (and complementary) alternative, with potential benefits (increased offloading) and costs (reduced QoE and potential monetary incentives)[3].

6.2.3 Cost Model

The goal of an offloading mechanism is to minimize the cost of delivering a set of contents over time to different nodes. Hence, we need first to define a model for the costs involved in each phase of the "offloading on the edge" mechanism.

– ***Initial Placement Costs:*** C_{BH}**,** C_{BS}**.**

The content provider, at time $t_0 = 0$, places the content to some edge nodes (SCs and/or MNs). A content is placed to a SC through a backhaul (wired or wireless) transmission, and we denote this per placement cost as C_{BH}. A (possible) content placement to some MNs takes place through a macro-cell BS transmission. We denote this transmission cost, which mainly depends on the load/congestion of the BSs, as C_{BS}.

– ***Opportunistic Offloading Costs:*** C_{SC}**,** C_{D2D}**.**

During time $t \in (0, TTL]$, the holders (which are either SCs or MNs) deliver the content to any requester they meet. We consider different costs for a SC-MN and a MN-MN (or D2D) transmission: C_{SC} and C_{D2D}. The former cost depends on the operating cost (transmission, energy consumption) of an SC, whereas the latter might exist if a compensation (or reward) is given by operator to MNs for each content they offload.

– ***Delayed Delivery Cost:*** $C_{BS}^{(TTL)}$**.**

At time TTL, the cellular network sends through macro-cell BSs the content to every non-served requester. This cost relates both to the load of BS (as C_{BS}) and to a (possible) compensation to the MNs for a delayed delivery. We denote this (per transmission) cost as $C_{BS}^{(TTL)}$.

6.2.4 Content Dissemination Model and Assumptions

Let us assume a content item (e.g. a popular video file) and a set of MNs interested in it. The content provider, at time $t_0 = 0$, places the content to the caches of some SCs and/or MNs. If by an expiry time TTL (if any), some of the interested MNs have not met any edge node (SC or MN) with the content, they are served by a macro-cell BS[4].

For the ease of reference, we define the following sets of "edge nodes" that are involved in the offloading process:

Definition 17. *A* requester *of a content is a mobile node (MN) that (a) is interested in the content and (b) has not received it yet. We denote the set of requesters at time t as* $\mathcal{R}(t)$.

Definition 18. *A* holder *of a content is an edge node (SC or MN) that stores the content and will forward it to its requesters. We denote the set of holders at time t as* $\mathcal{H}(t)$.

[3]Clearly, such delays might not be acceptable for all applications. However, many applications are inherently delay-tolerant, e.g. software updates, file downloads, one way streaming (e.g. YouTube or Netflix). Moreover, users might be willing to accept small or larger delays, if appropriate incentives are provided, and delayed content delivery has already been considered in a number of contexts, e.g [48, 134] .

[4]In the mechanism we consider, the content is cached only at the initial time, $t_0 = 0$, and macro-cell BSs deliver it only at its expiry time , $t = TTL$. Although one could place contents during time $t \in (0, TTL)$ as well, it has been shown (for similar settings) that placing contents at times $t \in (0, TTL)$ leads to a sub-optimal performance [124, 139].

We further denote the number of requesters and holders as:

$$R(t) = |\mathcal{R}(t)| \quad \text{and} \quad H(t) = |\mathcal{H}(t)|$$

where $\mathcal{H}(t) = \mathcal{H}_{SC}(t) \cup \mathcal{H}_{MN}(t)$ and $H(t) = H_{SC}(t) + H_{MN}(t)$

To model the level of participation of MNs in the offloading mechanism, we make the following assumption.

Assumption 3 (Cooperation)**.** *A requester acts as a holder for a content it has received with probability $p_c \in [0, 1]$. The probability p_c is equal among all nodes and contents.*

The probability p_c captures either the chance a node to forward the content (e.g. it has enough resources at the time) or the percentage of nodes who are "contracted" to help[5].

Finally, since edge nodes can exchange data only when they come within transmission range, the offloading is heavily affected by these *meeting events* between nodes. To this end, we assume the the meeting/contact events between two nodes $\{i, j\}$, $i \in \mathcal{MN}$ and $j \in \mathcal{MN} \cup \mathcal{SC}$, follow the model of Def. 3.

6.3 Analysis

An operator, in order to optimize the offloading performance and cost, has to weigh its options and take decisions about: *how to deliver* each content (directly or through offloading), *how many copies* of a content should be placed to different edge nodes, *which contents to store* in the SC and/or MN caches when their capacity is limited, etc. To this end, in this section, we provide the analytical results that are needed when trying to answer these questions. Specifically, we provide simple, closed form expressions for the performance of the "offloading on the edge" mechanism (Section 6.3.1), and the costs it incurs (Section 6.3.2).

6.3.1 Content Dissemination

The performance of the "offloading on the edge" mechanism depends on how much traffic it can offload and/or how fast are contents delivered. To answer these questions, we calculate the two main (and most common) performance metrics, namely the *content delivery probability*, and *content delivery delay*.

First, we state the following Lemma, in which we use a mean field approximation and a resulting system of ODEs to approximate the number of holders and requesters over time.

Lemma 9. *The fluid-limit deterministic approximation for the expected number of holders ($H(t)$) and requesters ($R(t)$) at time t, is*

$$H(t) = H_0 \cdot \frac{(p_c \cdot R_0 + H_0) \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot t}}{p_c \cdot R_0 + H_0 \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot t}}$$

$$R(t) = R_0 \cdot \frac{p_c \cdot R_0 + H_0}{p_c \cdot R_0 + H_0 \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot t}}$$

[5]Here, we need to stress that the above assumption implies that MNs will never become holders of a content they are not interested in. Although there exist studies that assume that even not interested MNs might be willing to act as holders [17,124,139,141], we believe that incentive mechanisms for these cases are difficult to implement (e.g. a user easier accepts to forward a content it already has stored, than to retrieve, cache and forward a content it will never use). Nevertheless, our framework could be easily extended also for such cases.

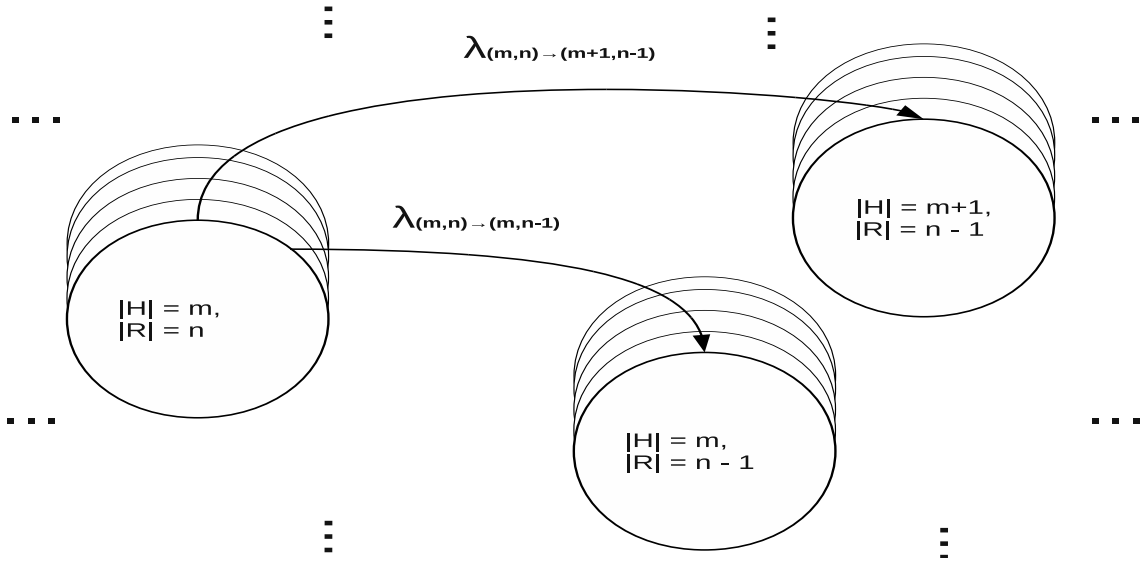


Figure 6.1: Content dissemination modeled by a Markov Chain.

where $H_0 = H(0^+)$ *and* $R_0 = R(0^+)$.

Proof. Having assumed Poisson meeting processes, we can model the dissemination of a content with a continuous Markov Chain, whose states correspond to the different sets of holders and requesters $\{\mathcal{H}, \mathcal{R}\}$. Fig. 6.1 shows a segment of this Markov Chain; we present the different states with equal number of holders ($|\mathcal{H}|$) and requesters ($|\mathcal{R}|$) under the same group, which can be described by the tuples $\{|\mathcal{H}|, |\mathcal{R}|\}$. To transition between states a *content delivery*, which takes place when a holder $i \in \mathcal{H}$ and a requester $j \in \mathcal{R}$ meet, is needed: (i) *Content delivery to cooperative node.* The next state is $\{|\mathcal{H}| = m + 1, |\mathcal{R}| = n - 1\}$ and the transition rate

$$\lambda_{(m,n)\to(m+1,n-1)} = p_c \cdot \sum_{i\in\mathcal{H}} \sum_{j\in\mathcal{R}} \lambda_{ij} \tag{6.1}$$

(ii) *Content delivery to non-cooperative node.* The next state is $\{|\mathcal{H}| = m, |\mathcal{R}| = n - 1\}$ and the transition rate

$$\lambda_{(m,n)\to(m,n-1)} = (1 - p_c) \cdot \sum_{i\in\mathcal{H}} \sum_{j\in\mathcal{R}} \lambda_{ij} \tag{6.2}$$

Statistics for the content dissemination process over time (e.g. distribution of $|\mathcal{H}(t)|$ or $|\mathcal{R}(t)|$), can be computed using the transition matrix of the Markov Chain of Fig. 6.1. However, this would render the problem analytically (and numerically, for large networks) intractable. To this end, we approach the problem with a mean field approximation of stochastic reaction models [75].

We first form the *drift equation* [75, Theorem 1.4.1] for the expected number of holders, $E\left[|\mathcal{H}(t)|\right] \equiv E\left[H(t)\right]$, as:

$$\frac{dE\left[H(t)\right]}{dt} = E\left[\lambda_{(m,n)\to(m+1,n-1)}\right] = p_c \cdot E\left[\sum_{i\in\mathcal{H}} \sum_{j\in\mathcal{R}} \lambda_{ij}\right]$$

The expectation in the right side of the drift equation is difficult to compute, as it requires the computation of the probabilities over the whole state space $\{\mathcal{H}, \mathcal{R}\}$. To this end, one can approximate $E[H(t)]$ with its deterministic equivalent $h(t)$. This approximation comes after neglecting the variability of $H(t)$ around its mean value and becomes more accurate for larger systems [75, Section 1.5].

Based on the deterministic approximation and since (a) the rates λ_{ij} are drawn independently from a distribution $f_\lambda(\lambda)$ with mean value μ_λ ($E[\lambda_{ij}] = \mu_\lambda$), and (b) the sum $\sum_{i\in\mathcal{H}}\sum_{j\in\mathcal{R}}\lambda_{ij}$ consists of $|\mathcal{H}|\cdot|\mathcal{R}|$ terms, we can write

$$E\left[\sum_{i\in\mathcal{H}}\sum_{j\in\mathcal{R}}\lambda_{ij}\right] \approx h(t)\cdot r(t)\cdot\mu_\lambda \tag{6.3}$$

The higher the number of terms in the above sum, and the less the heterogeneity of the meeting rates (i.e. the variance of $f_\lambda(\lambda)$), the more accurate the approximation in Eq. (6.3) is.

Substituting Eq. (6.3) in the drift equation (where $H(t)\to h(t)$), gives the ordinary differential equation (ODE) for $h(t)$[6]

$$\frac{dh(t)}{dt} = p_c\cdot h(t)\cdot r(t)\cdot\mu_\lambda \tag{6.4}$$

Proceeding similarly, the ODE for the deterministic approximation of the number of requesters ($R(t)\to r(t)$), is

$$\frac{dr(t)}{dt} = -h(t)\cdot r(t)\cdot\mu_\lambda \tag{6.5}$$

Finally, solving the system of the ODEs of Eq. (6.4) and Eq. (6.5), gives the expressions of Lemma 9. □

Based on Lemma 9 we, now, proceed to the calculation of the performance metrics. Let us consider a requester $i\in\mathcal{R}(0^+)$, and denote as T_i the time it receives the content. The probability this (random) requester to receive the content by a time t, i.e. $P\{T_i\le t\}$, is equal to the *percentage of offloaded contents* by time t. Hence, we can write

$$P\{T_i\le t\} = \frac{R_0 - R(t)}{R_0} = 1 - \frac{R(t)}{R_0} \tag{6.6}$$

Substituting the expression of Lemma 9 in Eq. (6.6), gives the following Result for the content delivery probability

Result 16 (**Delivery Probability**). *The probability a content to be delivered to a requester by time t is given by*

$$P\{T_d\le t\} = 1 - \frac{p_c\cdot R_0 + H_0}{p_c\cdot R_0 + H_0\cdot e^{\mu_\lambda\cdot(p_c\cdot R_0+H_0)\cdot t}}$$

where $H_0 = H(0^+)$ and $R_0 = R(0^+)$.

With respect to the average delay a requester experiences till it receives the content, we state the following Result (the proof is technical and can be found in Section 6.8.1). We derive expressions for two cases: (a) the content does not expire (i.e. $TTL\to\infty$), and (b) a macro-cell BS serves undelivered contents at time $t = TTL$. .

[6]Note the differences between $H(t)$ and $h(t)$: (a) $H(t)$ is integer, whereas $h(t)$ is a real number; (b) the drift equation for $H(t)$ contains expectations, while the respective ODE for $h(t)$ does not.

Result 17 (**Delivery Delay**). *The expected content delivery delay, under an expiry time* $TTL \in [0, \infty)$*, is given by*
– for $p_c > 0$*:*

$$E[T_d|TTL] = \frac{\ln\left(1 + \frac{p_c \cdot R_0 - e^{-\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot TTL}}{H_0 + p_c \cdot R_0 \cdot e^{-\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot TTL}}\right)}{\mu_\lambda \cdot p_c \cdot R_0}$$

– for $p_c = 0$*:*

$$E[T_d|TTL] = \frac{1 - e^{-\mu_\lambda \cdot H_0 \cdot TTL}}{\mu_\lambda \cdot H_0}$$

where $H_0 = H(0^+)$ *and* $R_0 = R(0^+)$.

6.3.2 Content Delivery Cost

Incorporating the offloading costs (Section 6.2.3) in our content dissemination model, and using the analytical results of Section 6.3.1, we calculate the cost of a single content delivery in Result 18. The expression we derive, gives the cost as a (simple) function of the system parameters (e.g. R_0, μ_λ) and the operator selected parameters (e.g. $H_{SC}(0)$, $H_{MN}(0)$), providing, thus, the necessary information for the evaluation and tuning of the "offloading on the edge" mechanism.

Result 18. *The cost of "offloading on the edge" a content is given by*

$$\begin{aligned} C =& C_{BH} \cdot H_{SC}(0) + C_{BS} \cdot H_{MN}(0) \\ &+ (C_{SC} \cdot q + C_{D2D} \cdot (1 - q)) \cdot R_0 \cdot P\{T_d \leq TTL\} \\ &+ C_{BS}^{(TTL)} \cdot R_0 \cdot (1 - P\{T_d \leq TTL\}) \end{aligned}$$

where $q = \frac{H_{SC}(0) \cdot \ln\left(\frac{H(TTL)}{H_0}\right)}{p_c \cdot (R_0 - R(TTL))}$*, and* $P\{T_d \leq TTL\}$*,* $H(TTL)$ *and* $R(TTL)$ *are given in Lemma 9 and Result 16.*

Proof.
– Initial Placement. The first two terms correspond to the initial placement phase: The cellular network operator, at time $t = 0$, places the content to $H_{SC}(0)$ SCs and $H_{MN}(0)$ MNs; in total ($H_0 = H_{SC}(0) + H_{MN}(0)$) holders. The costs per content placement are C_{BH} and C_{BS}, respectively.

– Opportunistic Offloading. During the opportunistic offloading phase, i.e. $t \in (0, TTL)$, the average number of requesters that receive the content by an edge node is $R_0 \cdot P\{T_d \leq TTL\}$. If we denote with q the percentage of requesters that receive the content by a SC, it is easy to see that the costs due to SC-MN and MN-MN content deliveries are

$$C_{SC} \cdot q \cdot R_0 \cdot P\{T_d \leq TTL\} \tag{6.7}$$

$$C_{D2D} \cdot (1 - q) \cdot R_0 \cdot P\{T_d \leq TTL\} \tag{6.8}$$

respectively.

To calculate the percentage q we proceed as following:

At first, the total number of requesters that receive the content by time TTL is

$$\#R_{tot} = R_0 - R(t) \tag{6.9}$$

Second, the total number of requesters that receive the content in the interval $(t, t+dt]$, $t \in (0, TTL)$ is

$$R(t) - R(t, t+dt) = -dR(t) \tag{6.10}$$

The probability that a content delivery that takes place in the interval in the interval $(t, t+dt]$ is due to a SC is equal to

$$\frac{H_{SC}(0)}{H(t)} \in [0, 1] \tag{6.11}$$

where $H_{SC}(0)$ is the number of SC holders (which does not change over time), and $H(t)$ the total number of holders at time t.

Therefore, the number of requesters that receive the content by an SC holder in the interval $(t, t+dt]$ is given by $-dR(t) \cdot \frac{H_{SC}(0)}{H(t)}$, and the total number of requesters that receive the content by an SC holder by time TTL is

$$\begin{aligned} \#R_{SC} &= \textstyle\int_0^{TTL} -dR(t) \cdot \frac{H_{SC}(0)}{H(t)} = \int_0^{TTL} -\frac{dR(t)}{dt} \cdot \frac{H_{SC}(0)}{H(t)} \cdot dt \\ &\overset{\text{Eq. (6.5)}}{=} \textstyle\int_0^{TTL} H(t) \cdot R(t) \cdot \mu_\lambda \cdot \frac{H_{SC}(0)}{H(t)} \cdot dt \\ &= \mu_\lambda \cdot H_{SC}(0) \textstyle\int_0^{TTL} R(t) \cdot dt \end{aligned} \tag{6.12}$$

Using the expression of Lemma 9 for $R(t)$ to calculate the above integral, we get

$$\begin{aligned} \#R_{SC} &= \frac{H_{SC}(0)}{p_c} \cdot \ln\left(\frac{(p_c \cdot R_0 + H_0) \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot TTL}}{p_c \cdot R_0 + H_0 \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot TTL}}\right) \\ &= \frac{H_{SC}(0)}{p_c} \cdot \ln\left(\frac{H(TTL)}{H_0}\right) \end{aligned} \tag{6.13}$$

where the last equality follows from the expression for $H(t)$ given in Lemma 9.

Now, q easily follows from Eq. (6.9) and Eq. (6.13)

$$q = \frac{\#R_{SC}}{\#R_{tot}} = \frac{H_{SC}(0)}{p_c} \cdot \frac{\ln\left(\frac{H(TTL)}{H_0}\right)}{R_0 - R(TTL)} \tag{6.14}$$

− Delayed Delivery. Finally, the average number of requesters that do not receive the content before its expiry time, is given by $R_0 \cdot (1 - P\{T_d \leq TTL\})$. Since, the cost of each content transmission at time $t = TTL$ is $C_{BS}^{(TTL)}$, the total cost of delayed delivery phase (last line of the expression in Result 18) follows easily. □

6.4 Applications: Cost Optimization

In a real scenario, the network operator would have to offload simultaneously many different contents. Using the results of the previous section, the average performance or the total cost over all the contents can be calculated easily, by evaluating them for each content separately and

then averaging or summing them, respectively. However, since some of the system parameters are controlled by the operator (e.g. H_0), they can be selected such that they lead to optimal performance. To this end, in this section, as an application of our analytical results, we study how offloading and caching can be designed in order to minimize the total cost.

Let us assume that the content provider has to deliver $M \geq 1$ contents to their requesters. We denote the set of the contents as $\mathcal{M}$ ($M = |\mathcal{M}|$). Since in a real scenario, not all contents are expected to be equally popular [32, 38, 87], nor tolerate equal delays, we denote the popularity (i.e. the number of initial requesters) and the expiry time of each content $\theta \in \mathcal{M}$ as R_0^θ and TTL^θ, respectively.

Under a given setting (i.e. with certain mobility, cooperation, traffic, etc., characteristics), what the cellular network can select, is the initial placement (*caching*) for each content $\theta \in \mathcal{M}$; namely, the number of SC and MN initial holders, $H_{SC}^\theta(0)$ and $H_{MN}^\theta(0)$, respectively (note that $H_0^\theta(0) \equiv H_{SC}^\theta(0) + H_{MN}^\theta(0)$). Additionally, it might be possible that the delay-tolerance of each content, TTL^θ, can be selected as well.

Therefore, if we denote as C^θ is the delivery cost of a content $\theta \in \mathcal{M}$ (which is given by Result 18), we can express the *total* cost optimization problem as

Problem 1.

$$\min_{\overline{H_{SC}},\, \overline{H_{MN}},\, \overline{TTL}} \ \left\{\sum_{\theta\in\mathcal{M}} C^\theta\right\}$$

$$\begin{aligned} s.t. \ \ \forall\theta \in \mathcal{M}: \ & 0 \leq H_{SC}^\theta(0) \leq N_{SC} \\ & 0 \leq H_{MN}^\theta(0) \leq R^\theta(0) \\ & T_{min} \leq TTL^\theta \leq T_{max} \\ and \qquad & \sum_{\theta\in\mathcal{M}} H_{SC}^\theta(0) \leq \sum_{i\in\mathcal{SC}} Q(i) \end{aligned}$$

where $\overline{H_{SC}}$, $\overline{H_{MN}}$ and $\overline{TTL}$ denote the vectors with components $H_{SC}^\theta(0)$, $H_{MN}^\theta(0)$ and TTL^θ ($\theta \in \mathcal{M}$), respectively, and $Q(i)$ is the caching capacity (in number of contents) of a SC node i.

Remark: Since MNs cache only contents in which they are interested in, we assume that their storage capacity is enough for all the contents of interest. Hence, storage capacity constraints for MN are not considered in Problem 1.

Since the costs C^θ are expressed as a function of the optimization variables (Result 18), well known numerical methods can be employed to solve Problem 1. Under certain scenarios, analytical solutions for Problem 1 can be found as well. In the remainder, we focus on two characteristics cases, which are analytically solvable, and provide useful insights for the system.

6.4.1 Offloading through SCs

We first consider the case where contents are offloaded only through SCs (i.e. when $p_c = 0$ and $H_{MN}^\theta(0) = 0$, or equivalently, $H_0^\theta = H_{SC}^\theta(0)$). This is the most common and feasible scenario considered in previous literature, since MNs are not required to share their contents, and thus incentive mechanisms are easier to implement. In this case and for[7] $C_{SC} < C_{BS}^{(TTL)}$ it can be proved that Problem 1 is convex and we compute the analytical solution in Result 19. For

[7]The "offloading on the edge" mechanism is meaningful if $C_{SC} < C_{BS}^{(TTL)}$, as in the opposite case, offloading would cost more than directly delivering from the macro-cell BSs.

notation simplicity, we consider equal expiry times $TTL^\theta = TTL,\ \forall\theta \in \mathcal{M}$, and cache sizes $Q(i) = Q,\ \forall i \in \mathcal{SC}$. However, Result 19 can be easily modified for different[8] TTL^θ and $Q(i)$ values.

Result 19. *Under a base scenario ($p_c = 0$, $H_{MN}(0) = 0$), the initial allocation $\overline{H_{SC}}$ that minimizes the total cost, is given by*

$$H_{SC}^\theta(0) = \begin{cases} N_{SC} & ,R^\theta(0) > U \\ \frac{1}{\gamma}\cdot\ln\left(\frac{1}{L}\cdot R^\theta(0)\right) & ,L \le R^\theta(0) \le U \\ 0 & ,R^\theta(0) < L \end{cases}$$

with $\gamma = \mu_\lambda \cdot TTL$, $L = \frac{1}{\gamma\cdot\Phi}\cdot\left(1+\frac{\lambda_0}{C_{BH}}\right)$, $U = L\cdot e^{\gamma\cdot N_{SC}}$, $\Phi = \frac{C_{BS}^{(TTL)}-C_{SC}}{C_{BH}}$, and

$$\lambda_0 = \inf\left\{\lambda_0 \ge 0 : \sum_{\theta\in\mathcal{M}} H_{SC}^\theta(0) \le \sum_{i\in\mathcal{SC}} Q(i)\right\}$$

Proof. Applying the method of Lagrange multipliers [4] to Problem 1, gives (for brevity we use the notation $H_0^\theta \equiv\ H_{SC}^\theta(0^+) = H_{SC}^\theta(0)$ and $R_0^\theta \equiv R^\theta(0^+) = R^\theta(0)$):

$$\nabla\left(\sum_{\theta\in\mathcal{M}} C^\theta\right) = \nabla\lambda_0\left(\sum_{i\in\mathcal{SC}} Q(i) - \sum_{\theta\in\mathcal{M}} H_0^\theta\right) + \nabla\sum_{\theta\in\mathcal{M}}\lambda_\theta\cdot H_0^\theta + \nabla\sum_{\theta\in\mathcal{M}}\mu_\theta\cdot(N_{SC}-H_0^\theta) \quad (6.15)$$

where $\lambda_0 \ge 0$ and $\lambda_\theta, \mu_\theta \ge 0, \forall\theta\in\mathcal{M}$ are the lagrangian multipliers.

Using the expression of Result 16 for the delivery probability, the offloading cost (Result 18) of a content θ, in a base scenario, can be written as

$$C^\theta = C_{BH}\cdot H_0^\theta + C_{SC}\cdot R_0^\theta + (C_{BS}^{(TTL)} - C_{SC})\cdot R_0^\theta\cdot e^{-\mu_\lambda\cdot H_0^\theta\cdot TTL} \quad (6.16)$$

Substituting C^θ from Eq. (6.16) to Eq. (6.15), the differentiation over H_0^θ gives

$$H_0^\theta = \frac{1}{\gamma}\cdot\left[\ln\left(\Phi\cdot\gamma\cdot R_0^\theta\right) - \ln\left(1+\frac{\lambda_0-\lambda_\theta+\mu_\theta}{C_{BH}}\right)\right] \quad (6.17)$$

The conditions for the lagrangian multipliers, i.e.

$$\lambda_\theta\cdot H_0^\theta = 0, \quad \text{and} \quad \mu_\theta\cdot(N_{SC}-H_0^\theta) = 0 \quad ,\forall\theta\in\mathcal{M}$$

imply that H_0^θ either

(a) is given by Eq. (6.17) and $\lambda_\theta = \mu_\theta = 0$, or

(b) is equal to N_{SC} and $\lambda_\theta = 0, \mu_\theta > 0$, or

(c) is equal to 0 and $\lambda_\theta > 0, \mu_\theta = 0$

[8] In particular, one has to substitute γ with $\gamma^\theta = \mu_\lambda\cdot TTL^\theta$ for each content. The expressions for $H_{SC}^\theta(0)$ remain the same, and only the expressions of L and U need to be modified.

From condition (a), we calculate the limits of the interval within which the optimal H_0^θ is given by Eq. (6.17). To find the lower limit, L, we set H_0^θ (Eq. (6.17) with $\lambda_\theta = \mu_\theta = 0$) equal to 0 and for the upper limit, U, equal to N_{SC}, which give

$$L = \frac{1}{\gamma \cdot \Phi} \cdot \left(1 + \frac{\lambda_0}{C_{BH}}\right) \tag{6.18a}$$

$$U = \frac{1}{\gamma \cdot \Phi} \cdot e^{\gamma \cdot N_{SC}} \cdot \left(1 + \frac{\lambda_0}{C_{BH}}\right) = L \cdot e^{\gamma \cdot N_{SC}} \tag{6.18b}$$

Combining Eq. (6.17) and Eqs. (6.18), we can express the optimal placement as

$$H_0^{\theta\,*} = \begin{cases} N_{SC} & , R_0^\theta > U \\ \frac{\ln\left(\gamma \cdot \Phi \cdot R_0^\theta\right) - \ln\left(1 + \frac{\lambda_0}{C_{BH}}\right)}{\gamma} & , L \leq R_0^\theta \leq U \\ 0 & , R_0^\theta < L \end{cases} \tag{6.19}$$

The only unknown parameter in Eq. (6.19) is λ_0 (since we expressed L and U as functions of λ_0). Lemma 10, which we state and prove in Appendix 6.8.2, suggests that the total cost, $\sum_{\theta \in \mathcal{M}} C^\theta$, is monotonically increasing with λ_0. Therefore, the optimal placement policy corresponds to the smaller *non-negative* value of λ_0 that satisfies the storage constraint, $\sum_{\theta \in \mathcal{M}} H_0^\theta \leq \sum_{i \in \mathcal{SC}} Q(i)$, and this proves the Result. □

In general, the value of the parameter λ_0 can be found (within some precision) with e.g. a binary search. Nevertheless, for a large number of contents, and given their popularity distribution, its value can be directly calculated using the Corollary 4, which follows after substituting the expression of Result 19 and the popularity density function in the storage constraint $\sum_{\theta \in \mathcal{M}} H_{SC}^\theta(0) = \sum_{i \in \mathcal{SC}} Q(i)$.

Corollary 4. *Under a content popularity distribution $\rho(x)$, the parameter λ_0 in Result 19 is given by $\lambda_0 = \max\left\{0, \hat{\lambda_0}\right\}$, where $\hat{\lambda_0}$ is the (minimum) solution of*

$$\int_L^U \ln\left(\gamma \cdot \Phi \cdot x\right) \cdot \rho(x) dx - \ln\left(1 + \frac{\lambda_0}{C_{BH}}\right) \cdot \int_L^U \rho(x) dx + \gamma \cdot N_{SC} \cdot \int_U^\infty \rho(x) dx = \frac{\gamma \cdot N_{SC} \cdot Q}{M}$$

Result 19 reveals how resources should be allocated: (i) The optimal allocation is logarithmic in popularity, with either large or small caches. (ii) When capacity is limited, an extra factor (λ_0) is introduced, so that the *relative* allocation remains logarithmic, but the absolute allocation is reduced (normalized) as the number of contents increase, or total capacity decreases. (iii) Some low popularity contents might get no allocation, either because it does not help the offloading cost, or because there is not enough capacity for them.

Practical Example: Assume an urban area covered by $N_{BS} = 4$ macro-cell BSs and $N_{SC} = 100$ SCs. On average, in this area reside $N_{MN} = 10000$ users[9] with an average meeting rate $\mu_\lambda = 3.3 \cdot 10^{-5}$ meetings/sec (equal to this of the real mobility trace [58]). The cellular network has to deliver M contents (e.g. YouTube video files of an average size $10MB$ [38]) with

[9] Vodafone Germany reported an average number of 1700 users per cell (`http://mobilesociety.typepad.com/mobile_life/2009/06/base-station-numbers.html`). In an urban environment, users density is expected to be higher.

expiry time $TTL \approx 5min$ and popularity given by a bounded Pareto distribution in the interval $R_0 \in [10, 1000]$ with shape parameter $\alpha = 0.5$ [38]. The costs are[10] $C_{BS}^{(TTL)} = 10 \cdot C_{BH}$ and $C_{SC} \ll C_{BH}, C_{BS}^{(TTL)}$.

Substituting the given values, and taking the expectation over the popularity distribution, it follows that the necessary buffer size of a SC, $Q = \frac{E[H_0]}{N_{SC}} \cdot M \cdot L$, is approximately $1MB$ per content. This means that, even under very high traffic demand, the caching capacity of the SCs would be adequate such that the last constraint of Problem 1 is not violated; e.g. for $M = 100000$ (i.e. each user requests 10 videos per 5 minutes!), the needed capacity is $Q = 100GB$ (which is a feasible and relatively cheap investment).

6.4.2 Offloading through MNs

We now consider the case where offloading takes place only through MN-MN communication ($p_c > 0$) and *without* content storing on SCs (i.e. $H_{SC}(0) = 0$). A content is initially sent by the BSs to $H_{MN}(0)$ (out of $R(0)$) of its requesters, which start disseminating it to the other requesters. However, not all nodes might be willing to participate by acting as holders, which in our framework means that each node (including the initial nodes in which the content is placed) cooperates with probability p_c. Therefore, we can write

$$H_0 \equiv H_{MN}(0^+) = p_c \cdot H_{MN}(0)$$

Also, as defined in Lemma 9,

$$R_0 \equiv R(0^+) = R(0) - H_{MN}(0)$$

As in the previous case, we assume equal expiry times $TTL^\theta = TTL,\ \forall \theta \in \mathcal{M}$.

Result 20. *Under an opportunistic MN-MN scenario ($p_c > 0$, $H_{SC}(0) = 0$), the initial allocation $\overline{H_{MN}}$ that minimizes the total cost, is given by*

$$H_{MN}^\theta(0) = \begin{cases} R^\theta(0) & , R^\theta(0) \leq OPT^\theta \\ OPT^\theta & , 0 \leq OPT^\theta < R^\theta(0) \\ 0 & , OPT^\theta < 0 \end{cases}$$

where $OPT^\theta = \dfrac{R^\theta(0) \cdot \left(\sqrt{\Phi'} \cdot e^{\frac{1}{2}\gamma \cdot p_c \cdot R^\theta(0)} - 1\right)}{e^{\gamma \cdot p_c \cdot R^\theta(0)} - 1}$, *and* $\Phi' = \frac{C_{BS}^{(TTL)} - C_{D2D}}{C_{BS} - C_{D2D}}$ *and* $\gamma = \mu_\lambda \cdot TTL$.

Proof. The cost for offloading a content θ under an opportunistic MN-MN scenario, where $H_0^\theta = p_c \cdot H_{MN}^\theta(0)$ and $R_0^\theta = R(0)^\theta - H_{MN}^\theta(0)$, is (see Result 18)

$$C^\theta = C_{BS} \cdot H_{MN}^\theta(0) + \left(C_{D2D} - C_{BS}^{(TTL)}\right) \cdot (R^\theta(0) - H_{MN}^\theta(0)) \cdot P\{T_d \leq TTL\} + C_{BS}^{(TTL)} \cdot (R^\theta(0) - H_{MN}^\theta(0)) \quad (6.20)$$

[10] In general, the offloading costs incurred in each phase, might differ between areas, time periods and operators. Their absolute values are not available and/or are difficult to estimate. To this end, in this example, as well as in other numerical results, we use relative values inferred by some average values proposed in [66].

Similarly, for $H_0^\theta = p_c \cdot H_{MN}^\theta(0)$ and $R_0^\theta = R^\theta(0) - H_{MN}^\theta(0)$, the delivery probability $P\{T_d \leq TTL\}$ can be written as

$$P\{T_d \leq TTL\} = 1 - \frac{R^\theta(0)}{R^\theta(0) + H_{MN}^\theta(0) \cdot \left(e^{\gamma \cdot p_c \cdot R^\theta(0)} - 1\right)} \tag{6.21}$$

where $\gamma = \mu_\lambda \cdot TTL$.

Substituting Eq. (6.21) in Eq. (6.20), and taking the derivative over the initial number of transmissions $H_{MN}^\theta(0)$, gives

$$\frac{dC^\theta}{dH_{MN}^\theta(0)} = (C_{BS}^{(TTL)} - C_{D2D}) + \frac{(C_{D2D} - C_{BS}) \cdot (R^\theta(0))^2 \cdot e^{\gamma \cdot p_c \cdot R^\theta(0)}}{\left(R^\theta(0) + H_{MN}^\theta(0) \cdot (e^{\gamma \cdot p_c \cdot R^\theta(0)} - 1)\right)^2} \tag{6.22}$$

From Eq. (6.22) it follows that

$$\frac{dC^\theta}{dH_{MN}^\theta(0)} = \begin{cases} < 0 & , H_{MN}^\theta(0) < OPT^\theta \\ > 0 & , H_{MN}^\theta(0) > OPT^\theta \end{cases}$$

where

$$OPT^\theta = \frac{R^\theta(0) \cdot \left(\sqrt{\Phi'} \cdot e^{\frac{1}{2}\gamma \cdot p_c \cdot R^\theta(0)} - 1\right)}{e^{\gamma \cdot p_c \cdot R^\theta(0)} - 1} \tag{6.23}$$

Therefore, when $OPT^\theta \in [0, R^\theta(0)]$, the minimum cost is achieved for $H_{MN}^\theta(0) = OPT^\theta$. Otherwise, for $OPT^\theta \notin [0, R^\theta(0)]$, and since it must hold that $H_{MN}^\theta(0) \in [0, R^\theta(0)]$, the minimum cost is achieved for the largest or lowest possible values of $H_{MN}^\theta(0)$. □

Result 20 reveals how content storage should be delivered when offloading only through MNs is considered. As it can be seen, the initial allocation is much different than in the offloading through SCs case (see Result 19), and this is mainly due to the fact that some of the requesters get the content at the beginning.

6.5 Simulation Results

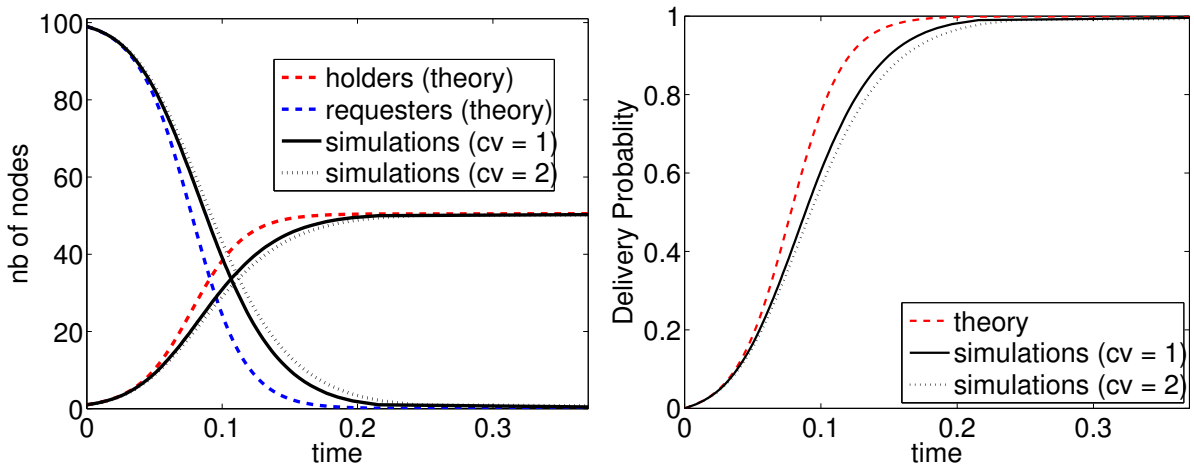
To validate our analysis, we compare the theoretical predictions against Monte Carlo simulations (Section 6.5.1). Then, we evaluate the cost efficiency of "offloading on the edge" in scenarios with realistic traffic demand patterns (Section 6.5.2).

6.5.1 Model Validation

6.5.1.1 Synthetic Scenarios

We first compare the theoretical results against Monte Carlo simulations on various synthetic scenarios. Synthetic simulations allow us to create a number of different scenarios with varying parameters.

We generate synthetic networks, conforming to the model of Section 6.2.4, as following:
(i) We choose a probability distribution $f_\lambda(\lambda)$ and for each pair $\{i, j\}$ we draw randomly a meeting rate λ_{ij}.
(ii) We create a sequence of contact events for every pair in the network with rate (Poisson



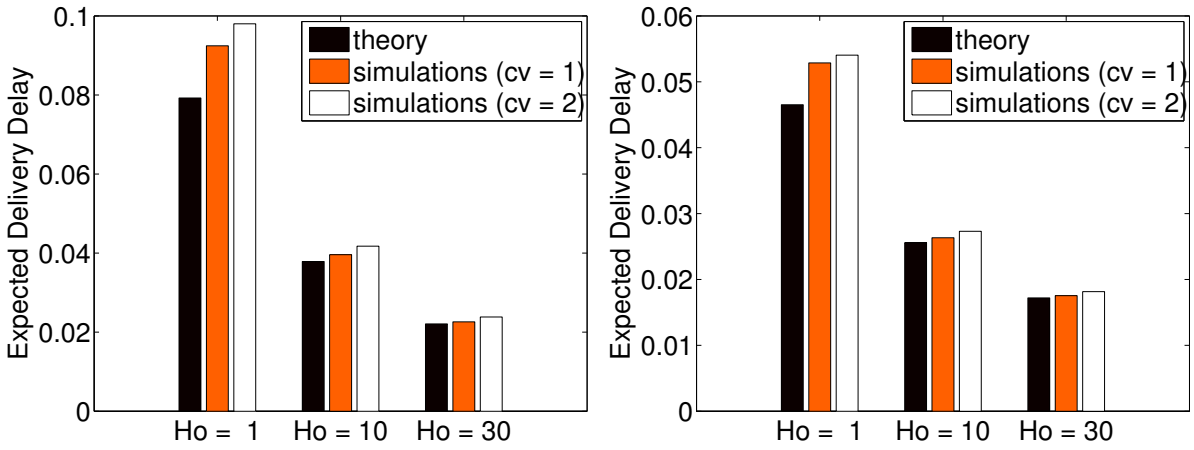
(a) $H_0 = 1$, $p_c = 0.5$ (b) $H_0 = 1$, $p_c = 0.5$

Figure 6.2: (a) Expected number of holders, $H(t)$, and requesters, $R(t)$, over time for generic scenarios with $R_0 = 100$, $H_{SC} = 0$; (b) shows the corresponding results for the delivery probability, i.e. $P\{T_d \leq TTL\}$, where TTL is the x-axis variable.

processes with rates λ_{ij}).
(iii) We select the content traffic parameters (R_0, H_0, p_c, $H_{SC}(0)$, $H_{MN}(0)$, N_{SC}), and we simulate a large number of content disseminations, choosing randomly each time the set of requesters and the set of holders (note, however, that the set of holders depends also on the parameters $H_{SC}(0)$, $H_{MN}(0)$ and N_{SC}).

We have created many scenarios with different combinations of mobility ($f_\lambda(\lambda)$) and traffic (R_0, H_0, p_c, $H_{SC}(0)$, $H_{MN}(0)$, N_{SC}) characteristics. We present here a representative subset of them, which allow us demonstrate the accuracy of our predictions and their sensitivity when varying certain parameters. In the presented scenarios we create nodes mobility according to a gamma distribution $f_\lambda(\lambda)$ with mean value $\mu_\lambda = 1$ (i.e. normalized value) and variance σ_λ^2 (or, equivalently, coefficient of variation $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$) [109]. Gamma distributions allow us to capture different levels of mobility heterogeneity by varying the value of CV_λ.

Content Dissemination. In Fig. 6.2 we compare simulation results (average values over the different runs) of expected number of holders ($H(t)$) / requesters ($R(t)$) and content delivery probability $P\{T_d \leq TTL\}$ with the respective theoretical predictions (Lemma 9 and Result 16, respectively). Considering the same content traffic parameters, we simulated scenarios with moderate ($CV_\lambda = 1$) and high ($CV_\lambda = 2$) mobility variance, in order to show how mobility heterogeneity affects the accuracy of our predictions. It can be seen that our predictions become more accurate for lower mobility heterogeneity ($CV_\lambda = 1$). This is due to the mean field approximation of the transitions rates we used in the analysis (see Section 6.3.1). For scenarios with even lower mobility heterogeneity (e.g. $CV_\lambda = 0.5$ - not shown in the plots) the accuracy is even better. Additionally, we need to highlight that these results correspond to an initial allocation of only one holder ($H_0 = 1$), which is the *worst case* scenario (i.e. lowest accuracy of the mean field approximation, and, thus our predictions) among the ones with the given mobility and traffic (other than H_0) characteristics. In the same scenarios, when considering a few more initial holders, e.g. $H_0 = 10$, theoretical results achieve an almost exact prediction.



(a) $p_c = 0.5$ (b) $p_c = 1$

Figure 6.3: Expected delivery delay, $E[T_d]$, for various generic scenarios with $R_0 = 100$, $H_{SC} = 0$ and (a) $p_c = 0.5$, (b) $p_c = 1$.

Similar observations can be made in Fig. 6.3, where we compare the theoretically predicted delivery delays with the respective simulation results. The results in Fig. 6.3 are in accordance with the above observations, i.e. the predictions' accuracy increases for (a) lower CV_λ, and (b) higher number of initial holders H_0.

Offloading Cost. We finally present results that validate the cost optimization analysis of Section 6.4. Fig. 6.4 shows the incurred cost for the cellular network (y-axis) under different number of initial holders H_0 (x-axis) for various generic traffic scenarios. Different cooperation policies (top plots: $p_c = 1$, middle plots: $p_c = 0.5$, and bottom plots: $p_c = 0$) and expiry times TTL (or, equivalently, $\gamma = \mu_\lambda \cdot TTL$) are considered. It can be seen that our results accurately predict the content dissemination cost.

Some remarkable observations about the optimal initial allocation of holders that can be made in Fig. 6.4 (as well as in other scenarios we investigated) are the following: (i) In many cases, offloading on the edge can significantly reduce the cost of a content dissemination. For instance, in the scenario shown in Fig. 6.4 (bottom plot - bottom curve / black color), even without node cooperation ($p_c = 0$), offloading on the edge can reduce the cost 10 times, compared to the corresponding scenario without offloading (i.e. $C = 100$). (ii) An optimal initial allocation requires only a small number of (initial) storage resources, which in most of the cases we present is equal or less than 20% of the content requesters. (iii) The higher the allowed delay (i.e. expiry time TTL or parameter γ) is, the larger the gain the cellular network can have is. For example, consider the red line ($\gamma = 0.05$) in the bottom plot. Increasing $\times 10$ the value of TTL (black line - $\gamma = 0.5$) can reduce the cost (e.g. for $H_0 = 5$ which is close to the optimal allocation) almost 8 times.

6.5.1.2 Mobility Traces

Results of synthetic simulations demonstrate a significant accuracy of our predictions and verify the arguments used in the derivation of our results. In this section, we present results in more challenging scenarios, where node mobility characteristics depart from our model assumptions.

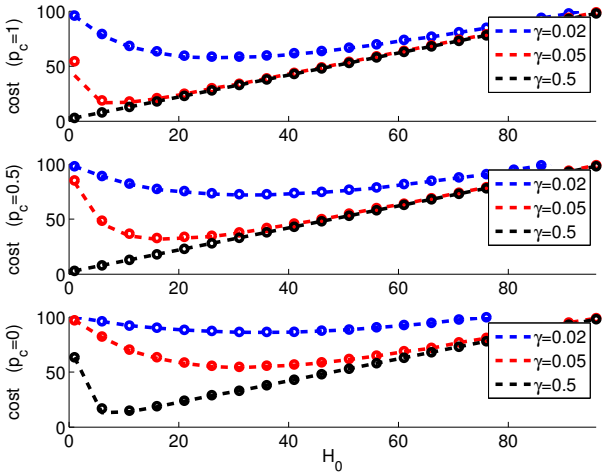


Figure 6.4: Single content offloading cost C (Lemma 18) under different number of initial holders (H_0, x-axis) for a synthetic mobility scenario with $R_0 = 100$, $H_{SC} = H_0$, and $C_{BH} = C_{BS}^{(TTL)} = 50 \cdot C_{SC}$. Dashed lines correspond to theoretical predictions and markers to simulation results. We denote $\gamma = \mu_\lambda \cdot TTL$.

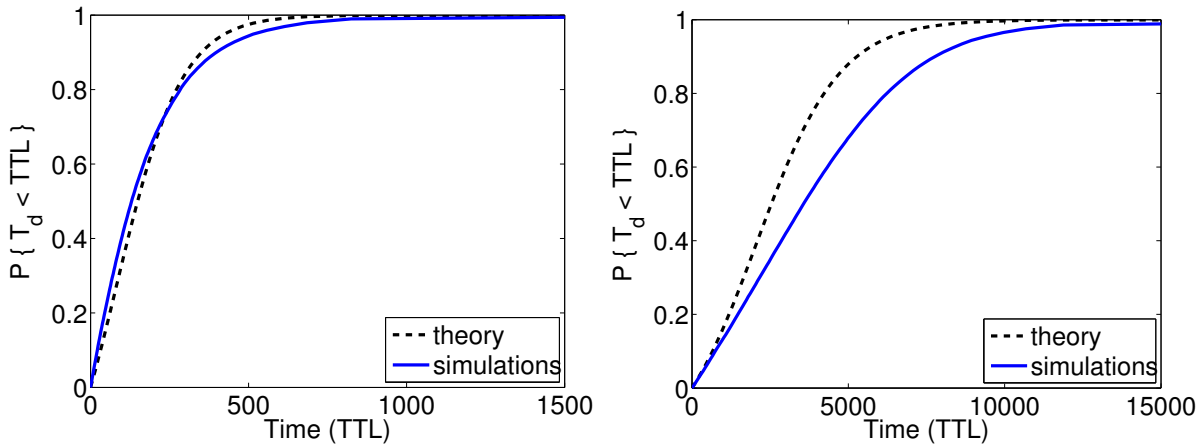
Specifically, we use the TVCM [57] and SLAW [77] mobility models, which have been shown to capture well real mobility patterns, like power-law flights [77], community structure [57], etc. The generated scenarios we present are

TVCM scenario: Mobile nodes move in a square area $1000m \times 1000m$, which contains three areas of interest (communities). Nodes move mainly inside their community (60% of the time) and leave it for a few short periods. Macro-cell BSs provide full coverage of the whole area, while 25 non-overlapping (placed on a grid) small-cell base stations (SCs), with a communication range of $100m$, provide further connectivity. Mobile nodes are equipped with *D2D* communication interfaces, for which we assume a range of $30m$.

SLAW scenario: A square area of edge length $2000m$ is simulated, where mobile nodes either move or remain static for a maximum time of $20min$ (the other mobility parameters are set as in the source code provided by [77]). Macro-cell BSs cover the whole area and coexist with 100 non-overlapping small-cells. Communication ranges are set as above.

In Fig. 6.5 we present the delivery probability $P\{T_d \leq TTL\}$, along with the theoretical prediction, for two content traffic scenarios in the TVCM (Fig. 6.5(a)) and SLAW (Fig. 6.5(b)) traces. Contents with popularity $R(0) = 50$ are initially cached to $H(0)$ edge nodes (half of which are MNs). The MNs' participation in offloading is set to $p_c = 0.5$. In the TVCM trace (Fig. 6.5(a)) it can be seen that the accuracy of our results is significant, despite the community structure of the network (which cannot be captured explicitly by our mobility assumptions). In the SLAW scenario (Fig. 6.5(b)), our results overestimate the delivery probability. However, note here that the number of holders in the SLAW scenario is smaller, and, thus, our approximation is expected to be less accurate. For scenarios with more initial holders the accuracy of the predictions increase (see e.g. Fig. 6.6(b), where the accuracy is higher for higher H_0 values).

Although in some points the theoretical performance metrics deviate considerably from simulations (e.g. 20%), the accuracy of the cost metrics (Lemma 18) is less affected. Fig. 6.6 shows the incurred cost for delivering a content to $R(0) = 30$ requesters (y-axis) under different number of initial holders H_0 (x-axis). Different initial placement policies $(H_{SC}(0), H_{MN}(0))$, levels of MNs participation (p_c), and expiry times TTL are considered. In the majority of scenarios our



(a) TVCM: $H(0) = 10$, $R(0) = 50$

(b) SLAW: $H(0) = 5$, $R(0) = 50$

Figure 6.5: Delivery probability $P\{T_d \leq TTL\}$ over time TTL (x-axis), for the (a) TVCM and (b) SLAW scenarios with $p_c = 0.5$ and $H_{SC}(0) = H_{MN}(0) = \frac{H(0)}{2}$.

results accurately predict the offloading cost. Yet, even in the case where the predictions are less accurate (e.g. in Fig. 6.6(b) for $\mu_\lambda \cdot TTL = 0.05$), they can still capture the actual optimal initial allocation regimes.

6.5.2 Performance Evaluation

After validating our analysis, we now investigate the cost efficiency of the "offloading on the edge" mechanism in a realistic traffic scenario. We present results that demonstrate the effect of different system factors, and provide useful conclusions for cellular network operators.

The parameters of the scenario we consider are the following:

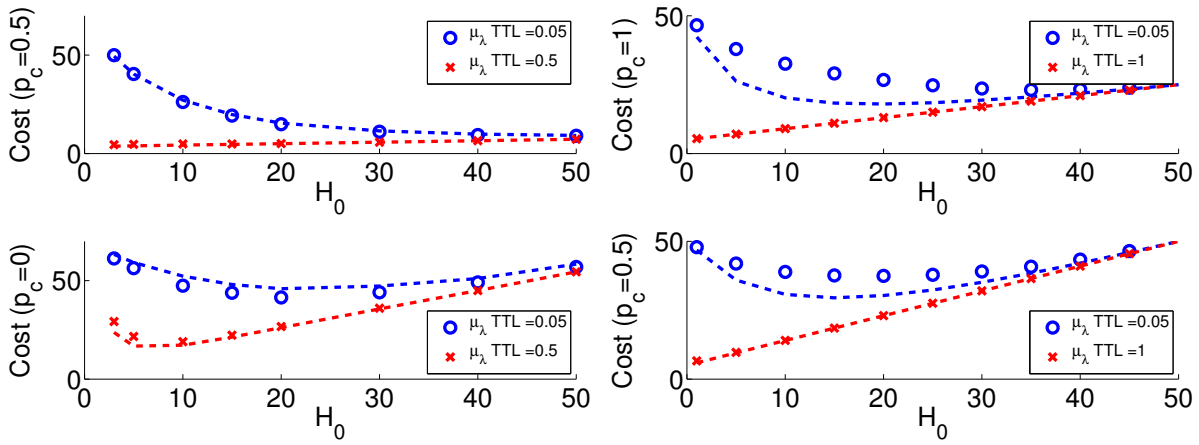
– *Popularity:* Content popularity has been shown to follow a power-law distribution [32,38,87]. Thus, we draw the popularity of each content from a bounded-Pareto distribution ($R_0 \in [1, 100]$) with shape parameter $\alpha = 0.5$ [38].

– *Traffic Intensity:* Mobile operators do not release real mobile traffic data. To this end, and since traffic demand is directly related to the number of mobile users that reside in an area, we infer traffic patterns from an available dataset of the Gowalla *location-based social network*. The Gowalla dataset [54] contains information (logs of position and time) of user checkins (through their mobile devices) in different venues. In the scenarios we present, we create different number of contents during a $24h$ time interval. The number of contents M is proportional to the number of mobile users that checked-in a certain area (we selected the most popular venue) at the same time. The maximum number of concurrent contents is $M = 200$.

– *Delay Tolerance:* We set equal expiry times TTL for each content, and we consider different sets of scenarios with low ($TTL = 5min$), moderate ($TTL = 25min$), and high ($TTL = 60min$) delay tolerance.

– *Costs:* The relative costs are set $C_{BS} = C_{BS}^{(TTL)} = 10 \cdot C_{BH} = 20 \cdot C_{SC} = 20 \cdot C_{D2D}$, values selected based on some data presented in [66].

– *Node Mobility:* We use the TVCM mobility scenario presented in the previous section.



(a) TVCM ($H_{SC}(0) = H_{MN}(0)$) (b) SLAW ($H_{SC}(0) = 0,$)

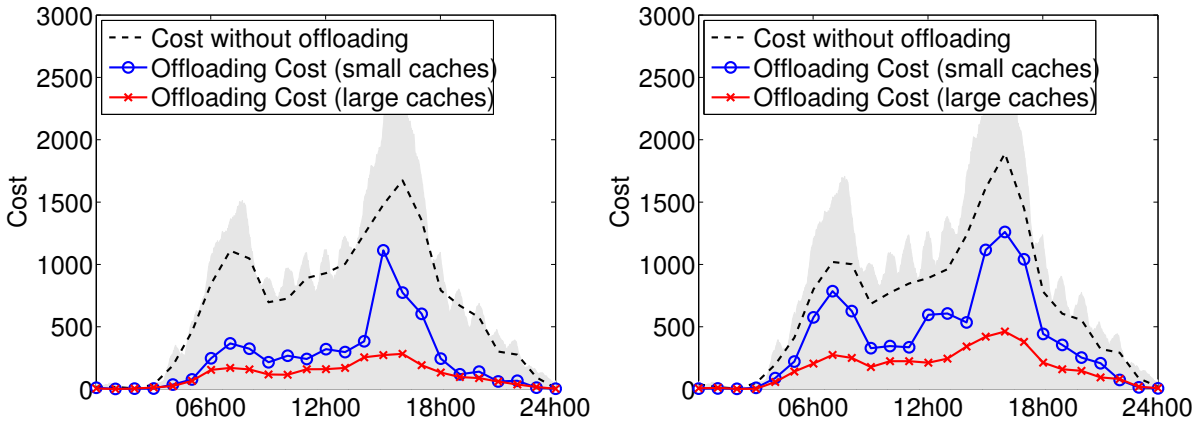
Figure 6.6: Offloading cost (y-axis) vs number of initial holders (H_0, x-axis). Dashed lines correspond to theoretical predictions and markers to simulation results. Transmission costs are: (a) $C_{BS}^{(TTL)} = 10 \cdot C_{BH} = 10 \cdot C_{BS} = 20 \cdot C_{SC} = 20 \cdot C_{D2D}$ (top plot) and $C_{BS}^{(TTL)} = C_{BH} = C_{BS} = 10 \cdot C_{SC}$ (bottom plot); (b) $C_{BS}^{(TTL)} = 2 \cdot C_{BS} = 10 \cdot C_{D2D}$.

Offloading through SCs

We first consider the case of offloading through SCs. We simulate two sets of scenarios with small ($Q = 5$) or large ($Q = 200$) caches. We choose the optimal initial caching policy of Result 19.

In Fig. 6.7 we present the total offloading cost (marked lines) incurred for the cellular network operator over different times of the day. The gray area shows the intensity of mobile users that reside in the considered area. The dashed line denotes traffic demand over time, or equivalently, the cost when content delivery *without* offloading is considered.

Some interesting observations that follow from Fig. 6.7 are:
(i) Under the optimal caching policy, "offloading on the edge" can significantly reduce the cost of content delivery, up to an order of magnitude, or even more in some cases.
(ii) The "offloading on the edge" cost changes over time much smoother than traffic demand. In particular, for large caches (cross/red line), the offloading cost curve is almost flat, despite the large peaks in traffic demand. In cellular networks, such temporal variations of the traffic intensity is an important issue, since operators are required to over-provision the network capacity (high CAPEX costs) [48]. As we show, "offloading on the edge" can amortize these costs. Even under higher transmission costs C_{BH}, C_{SC} than these we assumed, although the operating cost (OPEX) increases, the cost curve remains smooth, reducing thus a need for over-provisioning.
(iii) Large caching capacity has as a result a smoother cost curve (cross/red vs circle/blue curves). This is a positive message for operators, because to equip SCs with large enough caches is both feasible and inexpensive, as discussed in the example scenario of Section 6.4.1.
(iv) Comparing Fig. 6.7(a) and Fig. 6.7(b), we see that the tolerated delay has also a significant effect on the smoothness of the cost curve (higher TTL values lead to smaller variations). This implies that an alternative way of avoiding the over-provision cost (CAPEX), is to give incentives (OPEX) to users for accepting delayed content. Such solutions have been previously considered, e.g. [48], however, our framework allows an easy investigation of their effects (due to the closed-form results) and an analytic approach of pricing policies, etc.



(a) $TTL = 25min$ (b) $TTL = 5min$

Figure 6.7: Traffic demand and offloading cost over a $24h$ period.

Offloading through MNs

Now, we evaluate the performance of offloading through MNs. We simulate scenarios with different levels of node cooperation p_c. We choose the optimal initial content placement policy of Result 20.

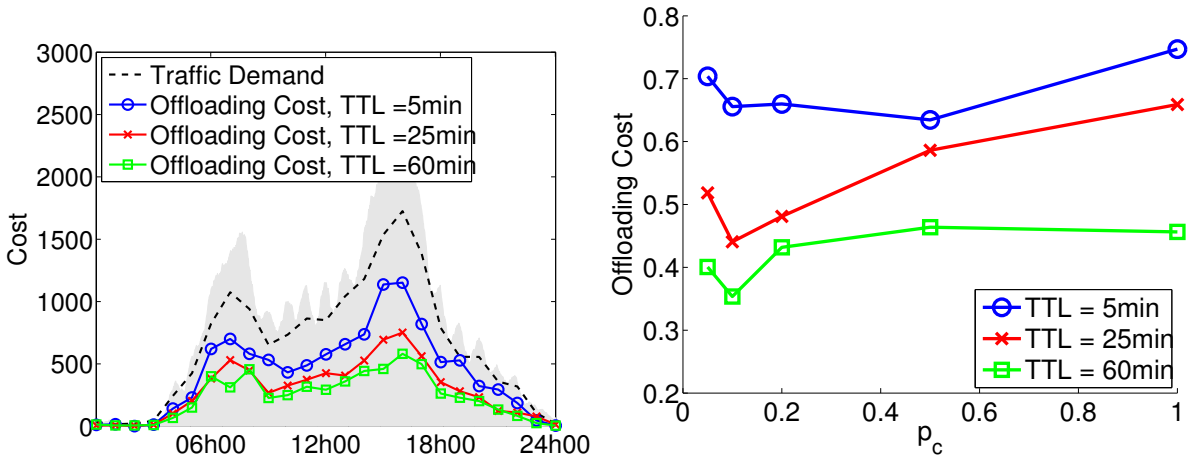
In Fig. 6.8(a) we present the total offloading cost (marked lines) incurred for the cellular network operator over different times of the day. We simulate three scenarios with low, moderate and high delay tolerance ($TTL = 5, 25, 60min$), and 10% of user cooperation in offloading ($p_c = 0.1$). Similarly to the offloading through SCs case (see e.g. Fig. 6.7), for higher TTL values, the cost decreases and its variations are smoother. However, it can be seen that improvement between the scenarios with $TTL = 25min$ and $TTL = 60min$ is not significant. This has an important implication for the system: Although increasing the delay tolerance is beneficial for the operator, after a point or gradually (depending on the scenario), the effects of this improvement become negligible. Bearing in mind that user satisfaction decreases with TTL indicates that there is a tradeoff, which should be carefully assessed by the system designer or considered for further optimization.

In Fig. 6.8(a) we show how the total offloading cost over a day period (normalized to the respective cost without offloading) changes with p_c. It is evident that varying user cooperation does not have the same effects for different scenarios, and that the minimum total cost is achieved at different values of p_c. This introduces one extra dimension, which can be used for system optimization as well. Such optimization options (with respect to TTL, p_c, etc.) could lead to interesting conclusions, and we believe it would be useful to consider them in future research.

6.6 Related Work

In this section we discuss works that are closer to ours, rather than studies which do not consider caching and/or delay tolerant delivery, and which are mainly based on pure infrastructure architectures, e.g. with WiFi access points [78] or small-cell base stations [3, 21], or on the D2D paradigm [5].

Mobile data offloading through MSNs and epidemic content dissemination is studied in [17,



(a) Offloading Cost vs. Time

(b) Offloading Cost vs. p_c

Figure 6.8: (a) Traffic demand and offloading cost over a $24h$ period. User cooperation is 10%. (b) Total offloading cost over a $24h$ period, normalized to the total cost without offloading.

124,139,141]. In the setting of [141], copies of a content are distributed through the infrastructure to a subset of mobile nodes, which then start propagating them epidemically. The performance of different content "pushing" techniques (e.g. slow/fast start) is investigated through simulations on a real vehicular mobility trace. Analytical approaches for pushing techniques can be found in [124,139], which study the optimal selection of the number of initial and final content pushes. [124] models the content dissemination as a control system and proposes an adaptive algorithm, *HYPE*, which aims to minimize the load of the cellular network by using real time measurements. On the other hand, [139] uses a fluid limit approximation and focuses on the cost optimization problem. Finally, [17] takes into account fairness among different contents/nodes, and derives schedulers that maximize the throughput, under given mobility and wireless channel conditions. These studies, in contrast to our framework, assume that *every* user is willing to offload contents, even if they are *not of her interest*. Difficulties in devising incentive mechanisms or limitations of device capabilities, might render such settings unrealistic.

To this end, [85] considers a limited number of (designated) holders. [85] proposes centralized algorithms for selecting the best set of available holders, in order to minimize the traffic load served by the infrastructure. Our paper extends this work, by introducing generic offloading costs and policies, and deriving insightful, closed-form results for the optimal caching.

Finally, [39] proposes caching in femto-cells and user devices, in a different setting than ours, where users communicate with several holders simultaneously. D2D communication is controlled by a macro-cell BS, which is aware of the status of caches, location of users, and channel state information between them. The objective of the paper is to decide which files should be stored and on which helper node, a problem that is shown to be *NP-hard*. This problem is formally presented, studied in more detail, and extended for coded contents in [126].

6.7 Conclusion

In this work we studied "offloading on the edge", a mechanism that employs edge nodes (SCs and/or MNs) to opportunistically offload popular content. We built a model that can capture heterogeneous traffic demand, user cooperation and mobility characteristics, and describe generic caching and offloading policies. Based on our model, we derived closed-form expressions for predicting the offloading performance. These allowed us to analytically study the cost optimization problem, and provide results that shed light on how caching policies should be designed. Realistic simulations verified the insights that stem from our analysis, and led to useful conclusions.

Our closed-form expressions reveal how and to what extent each system parameter affects performance and cost. Thus, they could be easily applied to sensitivity analysis, network planning and dimensioning, or design of pricing strategies; issues that have recently attracted a lot of attention from network operators, who seek novel solutions to alleviate the effects of the rapidly growing traffic demand.

6.8 Appendix: Supplementary Theoretical Results and Proofs

6.8.1 Proof of Result 17

Proof. The probability a content to be delivered in the time interval $[t, t+dt)$ is given by

$$P\{T_d \in [t, t+dt)\} = \frac{dP\{T_d \leq t\}}{dt} \cdot dt \tag{6.24}$$

Since a requester gets the content at time $t = TTL$ from a BS, if it has not received it earlier, we can write for the expected delay

$$\begin{aligned} E[T_i|TTL] &= TTL \cdot (1 - P\{T_d \leq TTL\}) + \int_0^{TTL} t \cdot P\{T_d \in [t, t+dt)\} \\ &= TTL \cdot (1 - P\{T_d \leq TTL\}) + \int_0^{TTL} t \cdot \frac{dP\{T_d \leq t\}}{dt} \cdot dt \end{aligned} \tag{6.25}$$

where the last equality follows from Eq. (6.24).

Using the expression of Result 16, we first compute the derivative $\frac{dP\{T_d \leq t\}}{dt}$, and, then, the integral in Eq. (6.25), and we get

$$\begin{aligned} E[T_i|TTL] = TTL \cdot (1 - P\{T_d \leq TTL\}) + \frac{1}{p_c \cdot R_0} \cdot \left(\frac{TTL \cdot H_0 \cdot (p_c \cdot R_0 + H_0) \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot TTL}}{p_c \cdot R_0 + H_0 \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot TTL}} \right) \\ + \frac{1}{\mu_\lambda \cdot p_c \cdot R_0} \cdot \ln \left(\frac{p_c \cdot R_0 + H_0}{p_c \cdot R_0 + H_0 \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot TTL}} \right) \end{aligned}$$

Substituting the value of $P\{T_d \leq TTL\}$ from Result 16 in the above equation, after some

algebraic manipulations, we can successively get

$$\begin{aligned} E[T_i|TTL] &= \frac{TTL\cdot(p_c\cdot R_0+H_0)}{p_c\cdot R_0}+\frac{1}{\mu_\lambda\cdot p_c\cdot R_0}\cdot\ln\left(\frac{p_c\cdot R_0+H_0}{p_c\cdot R_0+H_0\cdot e^{\mu_\lambda\cdot(p_c\cdot R_0+H_0)\cdot TTL}}\right)\\ &= \frac{1}{\mu_\lambda\cdot p_c\cdot R_0}\cdot\ln\left(\frac{(p_c\cdot R_0+H_0)\cdot e^{\mu_\lambda\cdot(p_c\cdot R_0+H_0)\cdot TTL}}{p_c\cdot R_0+H_0\cdot e^{\mu_\lambda\cdot(p_c\cdot R_0+H_0)\cdot TTL}}\right)\\ &= \frac{1}{\mu_\lambda\cdot p_c\cdot R_0}\cdot\ln\left(1+\frac{p_c\cdot R_0-e^{-\mu_\lambda\cdot(p_c\cdot R_0+H_0)\cdot TTL}}{H_0+p_c\cdot R_0\cdot e^{-\mu_\lambda\cdot(p_c\cdot R_0+H_0)\cdot TTL}}\right)\end{aligned}$$

which is the expression of Result 17 for $p_c>0$. The expression for $p_c=0$ follows after taking the limit ($p_c\to 0$) of the above expression. □

6.8.2 Lemma 10: Cost Monotonicity with λ_0

Lemma 10. *Under a content placement policy given by Eq.* (6.19)*, the derivative of the total cost,* $\sum_{\theta\in\mathcal{M}} C^\theta$*, with respect to* λ_0 *is*

$$\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{M}} C^\theta\right]=\frac{1}{\gamma}\cdot\left(1-\frac{1}{1+\frac{\lambda_0}{\Phi_1}}\right)\cdot|A|\geq 0$$

where $\mathcal{A}=\{\theta\in\mathcal{M}: L\leq R_0^\theta\leq U\}$.

Proof. From the conditions (b) and (c) (see, proof of Result 19), and similarly to Eqs. (6.18), we can express the multipliers λ_θ and μ_θ as a function of λ_0, as

$$\lambda_\theta=\begin{cases}\lambda_0+C_{BH}\left(1-\gamma\cdot\Phi\cdot R_0^\theta\right) & ,R_0^\theta<L\\ 0 & ,R_0^\theta\geq L\end{cases} \tag{6.26a}$$

$$\mu_\theta=\begin{cases}-\lambda_0-C_{BH}\left(1-\gamma\cdot\Phi\cdot e^{-\gamma\cdot N_{SC}R_0^\theta}\right) & ,R_0^\theta>U\\ 0 & ,R_0^\theta\leq U\end{cases} \tag{6.26b}$$

The cost of a single content dissemination, Eq. (6.16), under the content placement policy of Eq. (6.19), can be written as

$$\begin{aligned} C^\theta &= \frac{\Phi_1}{\gamma}\cdot\left[\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)-\ln\left(1+\frac{\lambda_0-\lambda_\theta+\mu_\theta}{\Phi_1}\right)\right]+\Phi_2\cdot R_0^\theta+(\Phi_3-\Phi_2)\cdot\frac{R_0^\theta\cdot\left(1+\frac{\lambda_0-\lambda_\theta+\mu_\theta}{\Phi_1}\right)}{\gamma\cdot\Phi\cdot R_0^\theta}\\ &= \frac{\Phi_1}{\gamma}\cdot\left[\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)-\ln\left(1+\frac{\lambda_0-\lambda_\theta+\mu_\theta}{\Phi_1}\right)\right]+\Phi_2\cdot R_0^\theta+\frac{\Phi_1}{\gamma}\cdot\left(1+\frac{\lambda_0-\lambda_\theta+\mu_\theta}{\Phi_1}\right)\end{aligned} \tag{6.27}$$

Taking its derivative, with respect to λ_0, gives

$$\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{M}} C^\theta\right]=-\frac{\Phi_1}{\gamma}\cdot\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{M}}\ln\left(1+\frac{\lambda_0-\lambda_\theta+\mu_\theta}{\Phi_1}\right)\right]+\frac{1}{\gamma}\cdot\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{M}}(\lambda_0-\lambda_\theta+\mu_\theta)\right] \tag{6.28}$$

because the terms including only the scenario parameters (R_0^θ, γ, and costs) do not depend on the selected resource allocation and, thus, on the parameter λ_0.

To calculate the derivatives appearing in the right side of Eq. (6.28), we use the definition of a derivative, i.e.

$$\frac{df(\lambda_0)}{d\lambda_0} = \lim_{d\lambda_0 \to 0} \frac{f(\lambda_0 + d\lambda_0) - f(\lambda_0)}{d\lambda_0} \tag{6.29}$$

and proceed as following:

We first define the sets

$$\mathcal{A} = \{\theta \in \mathcal{M} : L \leq R_0^\theta \leq U\} \tag{6.30a}$$
$$\mathcal{B} = \{\theta \in \mathcal{M} : R_0^\theta > U\} \tag{6.30b}$$
$$\mathcal{C} = \{\theta \in \mathcal{M} : R_0^\theta < L\} \tag{6.30c}$$

and, respectively, for $\lambda_0 \to \lambda_0 + d\lambda_0$, the sets

$$\mathcal{A}' = \{\theta \in \mathcal{M} : L + \Delta L \leq R_0^\theta \leq +\Delta U\} \tag{6.31a}$$
$$\mathcal{B}' = \{\theta \in \mathcal{M} : R_0^\theta > U + \Delta U\} \tag{6.31b}$$
$$\mathcal{C}' = \{\theta \in \mathcal{M} : R_0^\theta < L + \Delta L\} \tag{6.31c}$$

where we denoted

$$L + \Delta L = \frac{1}{\gamma \cdot \Phi} \cdot \left(1 + \frac{\lambda_0 + d\lambda_0}{C_{BH}}\right) = L + \frac{d\lambda_0}{\gamma \cdot C_{BH} \cdot \Phi} \tag{6.32a}$$
$$U + \Delta U = \frac{1}{\gamma \cdot \Phi} \cdot e^{\gamma \cdot N_{SC}} \cdot \left(1 + \frac{\lambda_0 + d\lambda_0}{C_{BH}}\right) = U + \frac{d\lambda_0}{\gamma \cdot C_{BH} \cdot \Phi} \cdot e^{\gamma \cdot N_{SC}} = (L + \Delta L) \cdot e^{\gamma \cdot N_{SC}} \tag{6.32b}$$

Regarding the first derivative term in Eq. (6.28), we proceed as following

$$\begin{aligned}
&\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{M}}\ln\left(1+\frac{\lambda_0-\lambda_\theta+\mu_\theta}{C_{BH}}\right)\right] \\
&\overset{\text{Eqs. (6.26)}}{=} \frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{A}}\ln\left(1+\frac{\lambda_0}{C_{BH}}\right)\right]+\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{B}}\left(\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)-\gamma\cdot N_{SC}\right)\right] \\
&\qquad\qquad +\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{C}}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)\right] \\
&= \frac{d}{d\lambda_0}\left[|\mathcal{A}|\ln\left(1+\frac{\lambda_0}{C_{BH}}\right)\right]+\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{B}}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)\right]-\gamma\cdot N_{SC}\cdot\frac{d|\mathcal{B}|}{d\lambda_0} \\
&\qquad\qquad +\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{C}}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)\right] \\
&= |\mathcal{A}|\cdot\frac{1}{C_{BH}}\cdot\frac{1}{1+\frac{\lambda_0}{C_{BH}}}+\ln\left(1+\frac{\lambda_0}{C_{BH}}\right)\cdot\frac{d|\mathcal{A}|}{d\lambda_0}+\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{B}}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)\right]-\gamma\cdot N_{SC}\cdot\frac{d|\mathcal{B}|}{d\lambda_0} \\
&\qquad\qquad +\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{C}}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)\right]
\end{aligned} \tag{6.33}$$

If we define as $\rho(x)$ the content popularity distribution, the derivatives in the above sum are calculated as following

$$\begin{aligned}
\frac{d|\mathcal{A}|}{d\lambda_0} &= \frac{|\mathcal{A}'|-|\mathcal{A}|}{d\lambda_0} \\
&= \frac{\int_{L+\Delta L}^{U+\Delta U} M\cdot\rho(x)dx-\int_L^U M\cdot\rho(x)dx}{d\lambda_0} \\
&= M\cdot\frac{\int_U^{U+\Delta U}\rho(x)dx-\int_L^{L+\Delta L}\rho(x)dx}{d\lambda_0} \\
&\approx M\cdot\frac{\rho(U)\cdot\Delta U-\rho(L)\cdot\Delta L}{d\lambda_0} \\
&\overset{\text{Eqs. (6.32)}}{=} M\cdot\frac{\rho(U)\cdot\Delta L\cdot e^{\gamma\cdot N_{SC}}-\rho(L)\cdot\Delta L}{d\lambda_0} \\
&= M\cdot\frac{\Delta L}{d\lambda_0}\cdot\left(\rho(U)\cdot e^{\gamma\cdot N_{SC}}-\rho(L)\right) \\
&\overset{\text{Eqs. (6.32)}}{=} \frac{M}{\gamma\cdot C_{BH}\cdot\Phi}\cdot\left(\rho(U)\cdot e^{\gamma\cdot N_{SC}}-\rho(L)\right)
\end{aligned} \tag{6.34a}$$

and, similarly,

$$\frac{d|\mathcal{B}|}{d\lambda_0}\approx -M\cdot\frac{e^{\gamma\cdot N_{SC}}}{\gamma\cdot C_{BH}\cdot\Phi}\cdot\rho(U) \tag{6.34b}$$

and

$$\begin{aligned}\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{B}}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)\right] &= \frac{\sum_{\theta\in\mathcal{B}'}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)-\sum_{\theta\in\mathcal{B}}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)}{d\lambda_0}\\ &= \frac{-\int_U^{U+\Delta U}\ln(\gamma\cdot\Phi\cdot x)\cdot M\cdot\rho(x)dx}{d\lambda_0}\\ &\approx -M\cdot\frac{\ln(\gamma\cdot\Phi\cdot U)\cdot\rho(U)\cdot\Delta U}{d\lambda_0}\\ &\overset{\text{Eqs. (6.32)}}{=} -M\cdot\frac{e^{\gamma\cdot N_{SC}}}{\gamma\cdot C_{BH}\cdot\Phi}\cdot\ln(\gamma\cdot\Phi\cdot U)\cdot\rho(U)\\ &\overset{\text{Eqs. (6.18)}}{=} -M\cdot\frac{e^{\gamma\cdot N_{SC}}}{\gamma\cdot C_{BH}\cdot\Phi}\cdot\rho(U)\cdot\left(\gamma\cdot N_{SC}+\left(1+\frac{\lambda_0}{C_{BH}}\right)\right)\end{aligned} \tag{6.34c}$$

and, similarly,

$$\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{C}}\ln\left(\gamma\cdot\Phi\cdot R_0^\theta\right)\right]\approx M\cdot\frac{1}{\gamma\cdot C_{BH}\cdot\Phi}\cdot\rho(L)\cdot\ln\left(1+\frac{\lambda_0}{C_{BH}}\right) \tag{6.34d}$$

Substituting Eqs. (6.34) in Eq. (6.33), gives

$$\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{M}}\ln\left(1+\frac{\lambda_0-\lambda_\theta+\mu_\theta}{\Phi_1}\right)\right]=|\mathcal{A}|\cdot\frac{1}{\Phi_1}\cdot\frac{1}{1+\frac{\lambda_0}{\Phi_1}} \tag{6.35}$$

Regarding the second derivative term in Eq. (6.28), we proceed as following

$$\begin{aligned}\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{M}}(\lambda_0-\lambda_\theta+\mu_\theta)\right] &\overset{\text{Eqs. (6.26)}}{=} \frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{A}}\lambda_0\right]+\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{B}}(\lambda_0+\mu_\theta)\right]+\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{C}}(\lambda_0-\lambda_\theta)\right]\\ &= \frac{d}{d\lambda_0}\left[\lambda_0\cdot|\mathcal{A}|\right]+\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{B}}\left(-C_{BH}+\gamma\cdot C_{BH}\cdot\Phi\cdot e^{-\gamma\cdot N_{SC}}\cdot R_0^\theta\right)\right]\\ &\quad+\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{C}}\left(-C_{BH}+\gamma\cdot C_{BH}\cdot\Phi\cdot R_0^\theta\right)\right]\\ &= |\mathcal{A}|+\lambda_0\cdot\frac{d|\mathcal{A}|}{d\lambda_0}-C_{BH}\cdot\frac{d|\mathcal{B}|}{d\lambda_0}+\gamma\cdot C_{BH}\cdot\Phi\cdot e^{-\gamma\cdot N_{SC}}\cdot\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{B}}R_0^\theta\right]\\ &\quad-C_{BH}\cdot\frac{d|\mathcal{C}|}{d\lambda_0}+\gamma\cdot C_{BH}\cdot\Phi\cdot\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{C}}R_0^\theta\right]\end{aligned} \tag{6.36}$$

Similarly as before, we get

$$\frac{d|\mathcal{C}|}{d\lambda_0}\approx M\cdot\frac{1}{\gamma\cdot C_{BH}\cdot\Phi}\cdot\rho(L) \tag{6.37a}$$

and

$$\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{B}} R_0^\theta\right] = \frac{-\int_U^{U+\Delta U} x\cdot M\cdot\rho(x)dx}{d\lambda_0} \approx -M\cdot\frac{U\cdot\rho(U)\cdot\Delta U}{d\lambda_0}$$
$$\overset{\text{Eqs. (6.32)}}{=} -M\cdot\frac{\Delta L}{d\lambda_0}\cdot L\cdot\rho(U)\cdot e^{2\cdot\gamma\cdot N_{SC}}$$
$$\overset{\text{Eqs. (6.18)}}{=} -M\cdot\frac{e^{\gamma\cdot N_{SC}}}{\gamma\cdot C_{BH}\cdot\Phi}\cdot\frac{1}{\gamma\cdot\Phi}\cdot\left(1+\frac{\lambda_0}{C_{BH}}\right)\rho(U) \tag{6.37b}$$

and, similarly,

$$\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{C}} R_0^\theta\right] \approx -M\cdot\frac{1}{\gamma\cdot C_{BH}\cdot\Phi}\cdot\frac{1}{\gamma\cdot\Phi}\cdot\left(1+\frac{\lambda_0}{C_{BH}}\right)\rho(L) \tag{6.37c}$$

Substituting Eqs. (6.37) in Eq. (6.36), gives

$$\frac{d}{d\lambda_0}\left[\sum_{\theta\in\mathcal{M}}(\lambda_0-\lambda_\theta+\mu_\theta)\right] = |\mathcal{A}| \tag{6.38}$$

Finally, substituting the expressions of Eq. (6.35) and Eq. (6.38) in Eq. (6.28), proves the Lemma. □

Chapter 7

Conclusions

Social characteristics of users affect the way they move and contact each other, their relationships, the way they communicate, their interests, etc. As a result, Mobile Social Networks comprising nodes with different social behaviors, exhibit also heterogeneity in characteristics that relate to performance of different communication mechanisms, like the node mobility or traffic patterns. A number of studies have studied this heterogeneity by analyzing real datasets (e.g. from experiments or other related networking paradigms) [25, 36, 56, 108] and identifying common patterns of nodes' mobility and relationships. Also, a lot of sophisticated communication protocols that exploit this heterogeneity in order to facilitate communication between users and increase the effectiveness of basic social-oblivious schemes, have been proposed, e.g. [29,59,60,98].

In general, depending on the characteristics of a network and the employed protocols, heterogeneity can have a positive or negative, significant or negligible impact on the communication performance (see e.g. [76, 146]). To predict and quantify this impact, it is common to use stochastic models and analysis. However, capturing the different social dimensions in MSNs, increases the complexity of models, and limits the applicability of results (e.g., results that are application-specific [20] or can be evaluated only with numerical methods [73]). Hence, it often becomes difficult to obtain intuition about how and to what extent heterogeneity affects performance.

To that end, in this thesis, we tried to analytically understand the effects of social characteristics under generic MSN settings. We proposed models that capture some important aspects of social heterogeneity (namely, mobility, selfishness, traffic demand, and interest patterns), but, at the same time, remain simple enough to allow tractable analysis and derivation of closed form results. The expressions for performance prediction we derived, are simple and require only the knowledge of a few network parameters. Thus, they can be easily used for fast evaluation, design and optimization of networking protocols, as well as for providing useful insights about the effects of social heterogeneity on performance.

Specifically, our contributions are summarized as following:

– We commenced our work with considering node mobility, due to its crucial role for communication in MSNs (Chapter 2). We defined a generic class of models that captures heterogeneity between contacting node pairs (*Heterogeneous Contact Networks*), as well as individual nodes (*Poisson* and *Configuration Model* sparse contact graphs). We derived results for the basic epidemic step delay, which can be used as the building blocks for predicting the performance in

a number of epidemic-based routing schemes. Our closed form expressions are direct functions of the contact rates and node degrees heterogeneity (CV_λ and CV_d, respectively), showing thus clearly the ways it affects performance.

A further utility of the *Heterogeneous Contact Network* model, is that it allows to incorporate in the analysis social characteristics that (i) relate to mobility and (ii) affect performance, e.g. as we did for *social selfishness* (Chapter 3) and *end-to-end traffic heterogeneity* (Chapter 4).

– Based on evidence for correlation between node mobility and social relationships, in Chapter 3, we proposed a framework that can describe social selfishness. Our analysis led to expressions that show how selfishness affects message delivery delay and delivery probability. Additionally, our model and results, can be applied to cases where node cooperation is determined by an external policy (related to social characteristics / mobility of nodes) rather than their willingness to relay messages. To this end, we used our framework to investigate if and how it can be possible to improve the performance - power consumption trade-off, by choosing a cooperation policy wisely.

– In Chapter 4, after showing that traffic patterns affect performance only jointly with mobility, we derived results for the performance of end-to-end communication mechanisms. We also shown how a positive correlation between pairwise traffic and mobility patterns, reduces the gap between direct transmission and relay-assisted protocols, and discussed the important implications this has for the design of routing protocols, as well as the feasibility of different applications for MSNs.

We then, turned our attention to the content-centric applications, in which, although communication is based on the same principles as in end-to-end communication (i.e. direct transmission or store-carry-forward), the *interest patterns* of users play a determinant role.

– We first (Chapter 5) modeled in a generic and application-independent way the interest patterns of users, namely content popularity and availability. Based on this model, we provided results for the performance prediction of content-centric mechanisms as a function of the content *popularity* and *availability*, and the node *mobility* statistics. Using our expressions (e.g. the simple bounds/approximations), one can obtain insights about how each of the three aforementioned factors can affect performance. To demonstrate this, we used them, in an example case, for optimizing the performance of mobile data offloading mechanisms by selecting the content allocation (which corresponds to content availability) policy.

– Prompted by the recent, large interest in offloading overloaded cellular networks, in Chapter 6, we further focused our content-centric communication analysis on the case of mobile data offloading. We built an analytical model that can describe offloading through *local data storage* on edge nodes (i.e. small-cells or users' portable devices) and *opportunistic communication*. We derived closed form expressions that predict the offloading performance and cost (for the operator), and depend only on (i) the number of users interested in a content (R_0), (ii) the number of edge nodes selected by the network to store the content (H_0), (iii) the users mobility (μ_λ) and cooperation (p_c) average statistics, and (iv) the content's delay tolerance (TTL). Using these expressions, it is possible for the operator to design and optimize the offloading system, e.g. by properly selecting the parameters H_0 or TTL for each content. Towards this direction, we provided some initial results for the optimal content placement under a given set of contents with known popularities (R_0).

Future Research

The main goal of initial studies in MSNs was to explore the possibilities and limitations of this new networking paradigm, where communication is opportunistic, message delivery is delay tolerant, relay nodes store and carry messages, etc. Later studies, as well as this thesis, investigate the effects of social heterogeneity in the communication performance. Since, till now, large scale MSNs have not been deployed in real world, these works consider generic settings and/or are based on assumptions (e.g. about the network characteristics and the communication schemes) that seem to be rational or are true in related networking environments (e.g. online social networks, ad-hoc networks, etc.).

However, the advances in mobile technology, the increase in the density of portable devices, the overload of cellular networks, and the emergence of new needs among consumers, betoken that the day of real MSNs deployment and their use in a regular base (maybe as frequently we use, e.g., mobile Internet or geo-location applications) is not far. Hence, it is expected to become soon clearer what the main trends and applications supported by MSNs would be.

As a result, more data and knowledge about the characteristics of MSNs will be available, and this will allow researchers to approach problems in a more realistic way. To this end, and given this knowledge, we believe that some important future research directions in Mobile Social Networking would (or should) be related to

- **Traffic Locality.** Location-based applications are very probable to become one of the major MSN applications. A factor indicating this is the increasing popularity of the corresponding location-based mobile applications. Second, the congestion of cellular networks, leads operators to consider alternative solutions for serving mobile data demand, and MSNs can act as a local traffic filter that will alleviate the congestion of the infrastructure. Therefore, we deem an investigation of the traffic demand patterns generated by such location-based applications quite important.

- **Service Load.** Portable devices, even phones, are rich in computing and software resources, enabling thus collaborative computing or mobile cloud computing. However, the patterns of the traffic generated in such applications (e.g. input/output of service jobs) might be much different than the traffic of pure communication applications (e.g. much larger files, more frequent data exchanges, etc.), and thus different research approaches would be needed. For instance, contact duration, which is usually neglected in existing models, might be determinant for the performance of applications where mainly large files are exchanged. Also, models for communication involving many concurrent jobs/packets, would become more accurate when incorporating queueing theoretical approaches.

- **Recommended Contents.** Existing content dissemination mechanisms for MSNs assume that a user is interested only in certain contents or categories of contents (e.g. podcasting, publish-subscribe applications). Nevertheless, recommendation systems (e.g. in web-applications) have shown that people often are satisfied with similar contents. Hence, an MSN mechanism providing also similar contents (when the requested are not available), can expedite content delivery but maybe decrease user satisfaction. This leads to a trade-off, which if manipulated properly, can be beneficial for the total network performance. Recommendation mechanisms (correlated with node mobility), user behaviors, and performance evaluation of such systems, compose an interesting direction in MSN research.

Chapter 8

Résumé [Français]

L'évolution récente des communications mobiles et la large diffusion des appareils mobiles "intelligents" ("smartphones") ont radicalement changé la façon dont nous communiquons. En particulier, les appareils mobiles permettent d'accéder à Internet, d'utiliser les réseaux sociaux en ligne et les réseaux basés sur la géolocalisation ainsi que de profiter de débits de données rapides. Par ailleurs, les appareils portables sont devenus puissants: ils supportent de multiples interfaces radio sans fil, utilisent de nombreux capteurs, offrent de larges espaces de stockage, etc. Toutes ces capacités ont permis l'avènement de nouveaux paradigmes de communications, de nouveaux services ainsi que de nouvelles applications.

La communication entre les utilisateurs via les réseaux cellulaires ou via Internet peut maintenant être complétée par la communication directe entre les appareils mobiles. Les utilisateurs peuvent directement échanger entre eux des données en utilisant seulement la communication sans fil locale (par exemple le Bluetooth ou le WiFi Direct), et ils peuvent former des réseaux mobiles avec les autres utilisateurs, en parallèle à un réseau cellulaire ou WiFi, ou même lorsque l'infrastructure est absent. Ces Réseaux Sociaux Mobiles (RSM) ont été envisagés afin d'améliorer les communications dans des environnements où les communications sont difficiles (e.g. zones rurales), ou bien afin d'améliorer les réseaux cellulaires, par exemple en déchargeant le réseau primaire.

Dans un RSM, un message peut être livré directement aux destinations quand ils se rencontrent avec le nœud-expéditeur(s) ("single-hop"). De manière alternative, le routage assisté de relais peut être employé : dans ce cas, les "nœuds-relais" stockent le message, le transportent lors de leurs déplacements, et peuvent également le transmettre à d'autres relais. De cette façon, le message peut finalement atteindre sa destination ("multi-hop"). Étant donné que la communication "mobile à mobile" n'a lieu que pendant les rencontres (ou "contacts") entre les nœuds, les performances des communications dans un RSM dépend fortement de la mobilité des nœuds. Aussi, la demande en terme de trafic (qui veut communiquer avec qui, ou qui s'intéresse à quoi) peut affecter de manière significative la performance de mécanismes de communication. En outre, de nombreuses études provenant de différentes disciplines (sociologie, réseaux opportunistes, médias sociaux, etc.) ont montré que les profils de mobilité et de trafic sont (a) largement hétérogènes et (b) corrélés avec les caractéristiques sociales des noeuds. En conséquence, les différents comportements sociaux des nœuds peuvent mener à un RSM très hétérogène.

Ainsi, l'objectif principal de cette thèse porte sur la compréhension, analytique, de l'impact

sur les RSM de l'hétérogénéité existante dans la mobilité, le trafic ou bien les caractéristiques sociales des nœuds. C'est donc dans cette optique que nous proposons de nouveaux modèles prenant en compte les aspects clés des caractéristiques des utilisateurs, et que nous analysons les performances des mécanismes de communication (e.g. protocoles de routage ou systèmes de distribution de contenu). Nous fournissons de nouveaux résultats concernant des aspects tels que l'égoïsme social et l'hétérogénéité du trafic, qui n'avaient pas été étudiés (analytiquement) jusqu'à présent. Enfin, sur la base de nos résultats, nous proposons des lignes directrices générales pour guider la conception de nouveaux protocoles et et de nouvelles applications au sein des RSM.

Les chapitres de la thèse sont organisées selon :

Chapitre 2 - Analyse du délai des schémas épidémiques dans les réseaux de contact hétérogènes, épars et denses

Chapitre 3 - Comprendre les effets d'égoïsme social

Chapitre 4 - Modélisation et analyse de l'hétérogénéité du trafic de la communication dans les RSM

Chapitre 5 - Effets des modes de popularité de contenu et de disponibilité de contenu

Chapitre 6 - "Offloading on the Edge": Analyse et optimisation du stockage local des données et de déchargement dans les HetNets

Dans la suite, nous donnons un bref résumé de chaque chapitre et soulignons nos principaux résultats.

8.1 Chapitre 2 – Analyse du délai de schémas épidémique au sein de réseaux de contact hétérogènes épars et denses.

Comme souligné précédemment, des noeuds ne peuvent communiquer s'ils sont à portée les uns des autres. Par conséquent, afin de pouvoir analyser les communications, il est nécessaire d'obtenir un modèle décrivant la façon dont des noeuds entrent en contact. Jusqu'à présent, afin de simplifier leur étude, les modèles de propagation épidémique se basaient sur des hypothèses relativement simples (*Random walk*, *Random waypoint*) où les mobilité des noeuds sont représentées par des processus stochastiques indépendants et identiquement distribués (c.f. [43, 49, 143]). Cependant, plusieurs études portant sur des observations réelles de mobilité [25, 36, 56, 108] ont révélé des conclusions sensiblement différentes. Deux aspects clés mis en avant par ces études sont (i) La fréquence avec laquelle deux noeuds entrent en contact peut varier de façon très importante en fonction des noeuds considérés. (ii) Beaucoup de paires de noeuds ne se rencontreront jamais. Ces conclusions remettent en question la précision et l'utilité des prédictions qui ont été faites à partir de modèles homogènes. Il est cependant importer de noter que dès que l'on s'écarte des hypothèses faites dans les modèles homogènes de mobilité, l'étude des communications entre noeuds mobiles devient très rapidement complexe et la zone d'application des résultats obtenus est souvent limitée [12, 36, 73, 76, 132].

Ces observations amènent la question suivante: *Est il possible d'obtenir des solutions de forme fermée qui soient à la fois précises et utiles à l'étude des performances de schémas de mobilité épidémique, en dépit de l'utilisation de modèle de mobilité plus génériques ?*

Afin de répondre à cette question, dans ce chapitre, nous considérons une importante classes de modèles de mobilité / contact possédant des fréquences de contact et des graphes de contact hétérogènes.

8.1.1 Réseaux de contact hétérogènes et schémas épidémique

Dans un premier temps nous faisons l'hypothèse que chaque paire de noeuds $\{i, j\}$ se rencontrent en suivant un processus aléatoire comprenant différentes fréquence de contacts, λ_{ij}, issus d'une distribution arbitraire, $f_\lambda(\lambda)$, dont la moyenne, μ_λ, et la variance, σ_λ^2 sont connues. Afin de décrire de tels réseaux nous utilisons les définitions suivantes:

Définition 8.1.1 (Réseau de contact)**.**
– Un réseau de contact, $\mathcal{N}$, est défini par un graphe $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ dont les sommets représentent les noeuds du réseau et l'existence d'une arête entre deux sommets représente un contact régulier entre les deux noeuds représentés par ces sommets.
– La séquence d'événements correspondant aux contacts entre chaque pair de noeud $\{i, j\}$ dont les sommets sont connectés par une arête ($\{i, j\} \in \mathcal{E}$) est représenté par un processus de point aléatoire avec une fréquence (taux) λ_{ij}.
– La durée d'un contact entre deux noeuds est négligeable comparée à la durée entre deux contacts. Cette durée est cependant suffisante pour que les transferts de données aient lieu.

Définition 8.1.2 (Réseau de contact hétérogène)**.**
Un réseau de contact hétérogène est défini comme un réseau de contact (Def. 8.1.1), où
– Les occurrences d'un contact entre deux noeuds $\{i, j\}$ suivent un processus de Poisson d'intensité λ_{ij}, ce qui signifie que les durées entre deux contacts sont indépendantes et distribuées de façon exponentielle avec une fréquence λ_{ij}.

– Les fréquences de contact, λ_{ij}, sont indépendants et issues d'une distribution arbitraire avec une densité de probabilité définie par la fonction: $f_\lambda(\lambda), \lambda \in [\lambda_{min}, \lambda_{max}] \subseteq (0, \infty)$, une moyenne finie μ_λ et une variance finie σ_λ^2 (coefficient de variation $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$).

8.1.1.1 Analyse asymptotique

A travers une analyse asymptotique du processus de propagation épidémique, nous obtenons le résultat suivant pour l'espérance du délai de propagation:

Théorème 8.1.1. *Lorsque la taille du réseau N augmente, l'erreur relative RE_k entre l'espérance du délai d'une étape $E[T_{k,k+1}]$ et la quantité $\frac{1}{k(N-k)\mu_\lambda}$ converge vers zéro*

$$\lim_{N\to\infty} RE_k = \lim_{N\to\infty} \frac{E[T_{k,k+1}] - \frac{1}{k(N-k)\mu_\lambda}}{E[T_{k,k+1}]} = 0$$

En d'autres termes, il est possible d'obtenir une estimation du délai à l'étape k avec une précision infinie de la façon suivante:

$$E[T_{k,k+1}] \approx \frac{1}{k(N-k)\mu_\lambda}$$

La Figure 8.1 présente les résultats obtenus à partir de plusieurs simulations qui démontrent la précision atteinte, grâce au résultat que nous avons obtenu, dans différent scénarios.

8.1.1.2 Réseaux de taille finie

Pour des réseaux de taille finies, nous utilisons la *Méthode Delta* [27, 103] afin d'obtenir les approximations suivantes qui nous permettent de prédire l'espérance du délai de propagation épidémique. Nous obtenons une expression à la fois simple, à forme fermée ne faisant appel qu'aux premier et second moments de la distribution de la fréquence de contact $f_\lambda(\lambda)$.

Résultat 8.1.1. *Dans un réseau de contact hétérogène (Def. 8.1.2) une approximation de l'espérance du délai d'étape peut être obtenue par:*

$$E[T_{k,k+1}] = \frac{1}{k(N-k)\mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{k(N-k)}\right) \tag{8.1}$$

8.1.1.3 Espérance du délai de protocoles de routage opportunistes

Afin de montrer la façon dont notre méthode peut être utilisée en pratique, nous estimons une expression à forme fermée pour le délai de différents protocoles. Les différentes expressions obtenues sont présentées dans la Table 8.1.

8.1.2 Graphes de contact épars

Nous étendons la classe des modèles de mobilité considérés en considérant de façon arbitraires des réseaux épars en autorisant une paire de nœud à ne jamais se rencontrer. Nous utilisons deux approches différentes. L'une basée sur les graphes de Poisson, l'autre basée sur les graphes de modèles de configuration.

8.1.2.1 Graphe de contact de Poisson

Nous étendons notre modèle de contact hétérogène (Def. 8.1.2) de la façon suivante:

Définition 8.1.3 (Réseau de contact de Poisson hétérogène)**.** *Pour chaque paire de noeud i et j nous pouvons observer les propriétés suivantes (i) Ces noeuds ont une probabilité $1-p_s$ de ne jamais se rencontrer, (ii) Ces noeuds ont une probabilité p_s de se rencontrer avec une fréquence λ_{ij}, selon le processus de contact défini dans la Def. 8.1.2.*

Nous prouvons, par la suite, que les prédictions de délais que nous obtenons peuvent être utilisé pour n'importe quel réseau épar modélisé par un graphe de Poisson en utilisant le corollaire suivant:

Corollaire 8.1.1. *Dans un réseau de contact de Poisson hétérogène (Def. 8.1.3), les résultats théoriques pour un réseau de contact hétérogène (Def. 8.1.2), sont obtenus en substituant les différents moments de la distribution de la fréquence de contact (μ_λ et σ^2_λ) avec les expressions*

$$\mu_{\lambda(p)} = p_s \cdot \mu_\lambda$$
$$\sigma^2_{\lambda(p)} = p_s \cdot \left[\sigma^2_\lambda + \mu^2_\lambda \cdot (1-p_s)\right]$$

8.1.2.2 Graphe de modèle de configuration

L'utilisation du modèle de configuration nous permet de modéliser des caractéristiques encore plus complexes des graphes du réseau de contact, plus particulièrement la distribution hétérogène du degré des différents noeuds.

Définition 8.1.4 (Modèle de configuration)**.** *Étant donné une taille de réseau N, et une distribution du degré de ses noeuds $\boldsymbol{p_d}$, ou bien de la séquence du degré de ses noeuds ($d_i,\ i = 1, ..., N$), le modèle de configuration génère des instances aléatoires de graphes $\mathcal{G}$, pour lesquels la distribution du degré de ses noeuds est $\boldsymbol{p_d}$. Les noeuds sont connectés de façon aléatoire et la probabilité pour deux noeuds i et j d'être connectés est proportionnelle au degré des noeuds i et j.*

Définition 8.1.5 (Réseau de contact de modèle de configuration hétérogène)**.**
– Étant donné une distribution de degrés $\boldsymbol{p_d}$, avec une moyenne μ_d et une variance σ^2_d (et $CV_d = \frac{\sigma_d}{\mu_d}$), un graphe de contact $\mathcal{G}$ est généré par un modèle de configuration.
– Chaque paire de noeuds i et j connecté par une arête, entre en contact avec une fréquence λ (identique pour les nœuds entrant en contact) en suivant un processus défini par la définition Def. 8.1.2.

Sous les hypothèse du modèle précédemment présenté (et conditionnellement à des fréquences de contact uniformes, c'est à dire $\lambda_{ij} = \lambda, \forall\{i,j\}$) il nous est tout de même possible d'obtenir une approximation du délai de propagation épidémique à travers une expression de forme fermée

$$E\left[T_{k,k+1}\right] = \frac{1}{\lambda} \cdot E\left[\frac{1}{D^{out}(k)}\right] \approx \frac{1}{\lambda \cdot \overline{D}^{out}(k)} \tag{8.2}$$

où

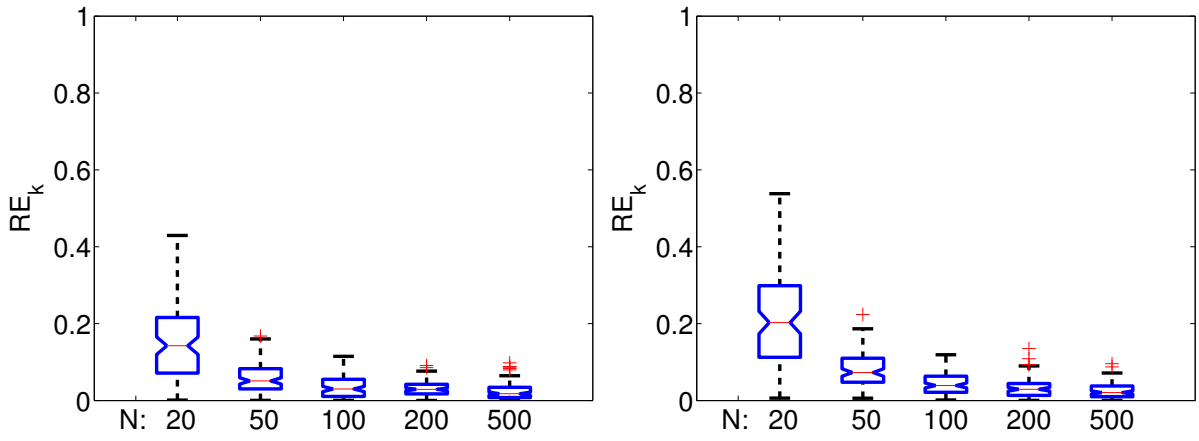
Résultat 8.1.2. *La moyenne du degré sortant à l'étape k, $D^{out}(k)$, est obtenue par l'approximation:*

$$\overline{D}^{out}(k) = (N-k)\mu_d \left[\left(\frac{N-k}{N-1}\right)^{CV_d^2} - \left(\frac{N-2}{N-1}\right)\left(\frac{N-k}{N-1}\right)^{2CV_d^2+1}\right] \tag{8.3}$$

8.1.3 Publications

Le travail présenté dans ce chapitre a donné lieu aux publications suivantes:

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Delay analysis of epidemic schemes in sparse and dense heterogeneous contact environments", Research Report RR-12-272, Eurecom, July 2012.*
- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Information diffusion in heterogeneous networks: The configuration model approach", Proc. 5th IEEE International Workshop on Network Science for Communication Networks (NetSciCom'13), co-located with IEEE INFOCOM 2013, 19 April 2013, Turin, Italy.*



(a) spreading step: 20% (b) spreading step: 70%

Figure 8.1: *Erreur relative d'étape* pour l'étape (a) $k = 0.2 \cdot N$ (le message a été propagé dans 20% du réseau) and (b) $k = 0.7 \cdot N$. Chaque boxplot correspond à une taille différente du réseau N (avec $\mu_\lambda = 1$ et $CV_\lambda = 1.5$). Les box-plots montrent la distribution de l'erreur relative d'étape RE_k estimée à partir de 100 instances pour chaque taille de réseau.

Table 8.1: Expressions d'estimation de l'espérance du délai pour différents protocoles de routage.

Epidemic	$E[T_D^{(epid)}] \approx \frac{1}{N \cdot \mu_\lambda} \cdot \left(\ln(N) + CV_\lambda^2 \cdot \frac{1.65 \cdot N + 2 \cdot \ln(N)}{N^2} \right)$
2-hop	$E[T_D^{(2-hop)}] = A_{N-1} \cdot \sum_{k=1}^{N-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+1} \cdot (N-k-1)!} \approx \frac{\sqrt{\frac{\pi}{2}}}{\sqrt{N} \cdot \mu_\lambda} \cdot \left(1 + \frac{CV_\lambda^2}{N} \right)$
SnW, L copies	$E[T_D^{(SnW)}] \leq A_{N-1} \cdot \sum_{k=1}^{L-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+1} \cdot (N-k-1)!} + (L \cdot A_{N-1} + A_L) \cdot \frac{(N-1)!}{(N-1)^L \cdot (N-L-1)!}$

$$\text{où } A_m = \frac{1}{m\mu_\lambda} \cdot \left[1 + \frac{CV_\lambda^2}{m} \right]$$

8.2 Chapitre 3 – Comprendre les effets d'égoïsme social.

Les modèles de mobilité définissent l'occurence de contacts entre les noeuds, en assumant que les échanges données ont lieu lors de ces contacts, il est alors possible d'évaluer différents algorithmes de routage et de transferts.

Cependant, la possibilité que les noeuds ne souhaite pas coopérer impacte fortement les techniques de dissémination des messages. C'est donc dans cette optique que, dans ce chapitre, nous étudions analytiquement l'effet de la coopération entre noeuds, ou égoïsme des noeuds, sur les MSNs.

8.2.1 Modèle d'égoïsme social

Dans un premier temps, nous étendons les études précédentes qui supposent des modèles d'égoïsmes uniformes, i.e. tous les noeuds ont la même réticence à coopérer.

Intuitivement la volonté d'un noeud de coopérer peut être reliée aux liens sociaux qui existent entre différents noeuds (i.e. des personnes qui se connaissent ont plus tendance à coopérer). De plus, des études ont montré que les liens sociaux ont un impact sur la fréquence avec laquelle des noeuds se rencontrent (deux personnes qui se connaissent ont plus tendance à se rencontrer que deux inconnus).

En combinant ces relations entre (i) l'égoïsme et les liens sociaux et (ii) les liens sociaux et les modèles de mobilités, il semble raisonnable de faire l'hypothése d'un modèle égoïste social où les noeuds utilisent l'opportunité d'un contact donné avec un probabilité $p_{ij} = p(\lambda_{ij})$, liée à la fréquence de contact entre les deux noeuds considérés $\{i, j\}$.

Afin de pouvoir capturer la plupart des comportement égoistes cités précédemment (voire plus) d'une façon à la fois simple et générique, nous décidons de modéliser cette volonté de transmettre un message (essentiellement, l'existence de contraintes associées affectant cette volonté) de manière probabiliste.

Plus précisément, nous proposons deux modèles d'égoïsme, qui correspondent aux comportements que l'on retrouve au sein d'un MSN.

Définition 8.2.1. *[égoïsme social: Type I] La probabilité pour un message d'être échangé lors d'un contact entre deux noeuds i et j dépend de leur fréquence de rencontre λ_{ij} et est décrite par la relation:*

$$p_{ij} = p^{(I)}(\lambda_{ij}), \qquad p_{ij} \in [0, 1] \tag{8.4}$$

Définition 8.2.2. *égoïsme social: Type II] Une paire de noeud i et j peut soit échanger un message lors de chaque rencontre avec la probabilité p_{ij} ou bien peut n'échanger aucun message avec la probabilité $1 - p_{ij}$. La probabilité p_{ij} dépend de la fréquence de rencontre de ces deux noeuds, i.e. λ_{ij}, et est décrite par la relation:*

$$p_{ij} = p^{(II)}(\lambda_{ij}), \qquad p_{ij} \in [0, 1] \tag{8.5}$$

8.2.2 Délai de livraison d'un message

En intégrant notre modèle d'égoïsme social au modèle du Chapitre 2, il nous est alors possible de combiner les effets de mobilités à ceux d'hétérogénéité égoïste. Il s'en suit donc la preuve du résultat suivant

Résultat 8.2.1. *L'espérance du délai de livraison d'un message dans un réseau de contact hétérogène peut être approximer par*

$$E[T_D] = \frac{c(N,L)}{\mu_\lambda^{eff.}}, \tag{8.6}$$

où $\mu_\lambda^{eff.} = E[\lambda \cdot p(\lambda)]$ *(*$p^{(I)}$ *ou* $p^{(II)}$ *représentant un égoïsme de Type I ou de Type II, respectivement) et* $c(N,L)$ *est un constant définie par la taille du réseau,* N*, le protocol de routage* $\mathcal{P}$ *et le nombre de copies de messages,* L*. Les valeurs de* $c(N,L)$ *sont données dans le Tableau 8.2 pour trois protocoles de routage connus.*

8.2.2.1 Validation

Afin de valider nos résultats, nous les comparons à ceux obtenus lors de simulations et, dans la Fig. 8.2 nous présentons, pour un réseau de $N = 100$ noeuds, la manière dont l'hétérogénéité mobile (i.e. $CV_\lambda = \frac{\sigma_\lambda}{\mu_\lambda}$) impacte le délai de livraison des message pour les différentes politiques d'égoïsme présentées dans le Tableau 8.3.

8.2.3 Compromis entre performance et consommation énergétique

Nous examinons à présent les régions de compromis entre performance et puissance qu'il nous est donné d'atteindre pour différentes politiques de coopération. Plus précisément, nous montrons que (i) lorsque l'on considére un classe intéressante de compromis Puissance vs Délai, des politiques "sociales" complexes ne peuvent pas atteindre de meilleures performances que celles d'une politique plus simpliste; alors que (ii) lorsque l'on considère des le compromis Puissance vs Probabilité de livraison les politiques de coopérations sociales peuvent, en effet, être optimisées.

8.2.3.1 Délai de livraison vs Consommation énergétique

Dans un premier temps, en utilisant un modèle simple d'injection de trafic, nous étudions le compromis entre délai de livraison de consommation énergétique. Nous prouvons que

Résultat 8.2.2. *La consommation énergétique moyenne d'un noeud est inversement proportionnelle au délai de livraison d'un message et est donnée par*

$$P = c_p \cdot \frac{1}{E[T_D]} \tag{8.7}$$

où $c_p = \frac{E_t \cdot N_f \cdot M \cdot c_t}{c}$.

il s'en suit donc que

Corollaire 8.2.1. *Dans un réseau de contact hétérogène, quelque soit la politique d'égoïsme utilisée, les régimes d'opération puissance-délai que l'on peut atteindre sont les mêmes. En d'autres termes, n'importe quel compromis entre puissance et délai qui peut être atteint par certain politique sociales d'égoïsme peut l'être aussi par un simple politique uniforme.*

Afin d'estimer la précison d'une telle prédiction, nous opérons plus simulations de type Monte Carlo et comparons les résultats obtenus expérimentalement avec ceux que nous obtenons

théoriquement. Les différentes simulations utilisent une politique d'égoïsme uniforme (Politique A). Puis nous lançons des simulations avec différentes politiques d'égoïsme non uniformes afin de pouvoir examiner si, en effet, la courbe de délai-puissance est la même. Comme il apparaît clairement sur la Fig. 8.3, nos résultats sont vérifiés par les simulations, c'est à dire en changeant les politiques d'égoïsme et leurs paramètres il est uniquement possible d'atteindre un déplacement sur la courbe théorique.

8.2.3.2 Probabilité de livraison vs Consommation énergétique

Dans un second temps, nous nous penchons sur le compromis entre probabilité de livraison et consommation énergétique qu'il est possible d'atteindre avec un mécanisme de partage de contenus. Nous prouvons que:

Résultat 8.2.3. *Dans un réseau de contact hétérogéne avec un politique d'égoïsme $p(\lambda)$, si N_A noeuds possèdent le contenu A, alors la probabilité pour un autre noeud d'accéder au contenu au bout d'un temps T est donné par*

$$P_A\{T\} = 1 - \left(E\left[e^{-\lambda \cdot p(\lambda) \cdot T}\right]\right)^{N_A} \tag{8.8}$$

où l'espérance est calculée sur f_λ.

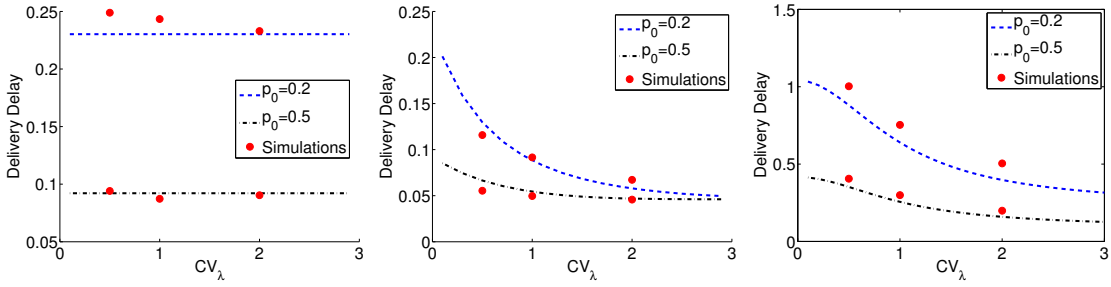
Dans ce cas, l'expression de la probabilité de livraison du contenu (Résultat 8.2.3) est liée au modèle de mobilité et à la politique d'égoïsme de manière non linéaire. Ainsi, cette non linéarité, indique qu'il est à présent possible de changer (et donc à terme d'améliorer la région puissance - performance que l'on peut atteindre).

Ce résultat est aussi observable sur la Fig. 8.4, où l'on compare la Politique A et C. Les résultats montrent que la Politique C offre de meilleurs résultats, dans le cadre de l'application de partage de contenus.

8.2.4 Publications liées à ces résultats

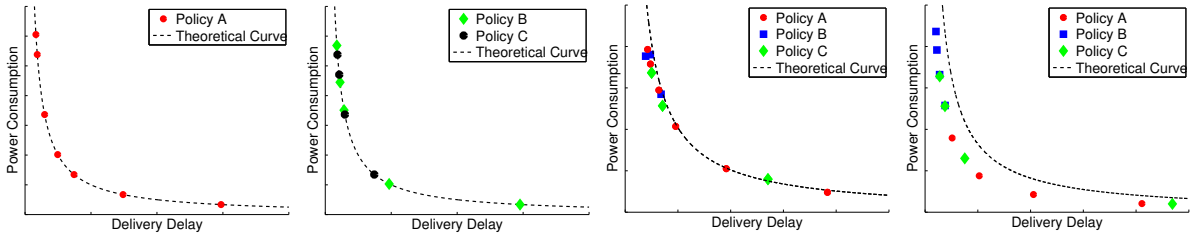
Le travail présent dans ce chapitre a donné lieu aux publications suivantes:

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Understanding the effects of social selfishness on the performance of heterogeneous opportunistic networks", Computer Communications, Elsevier, Volume 48, April 2014.*



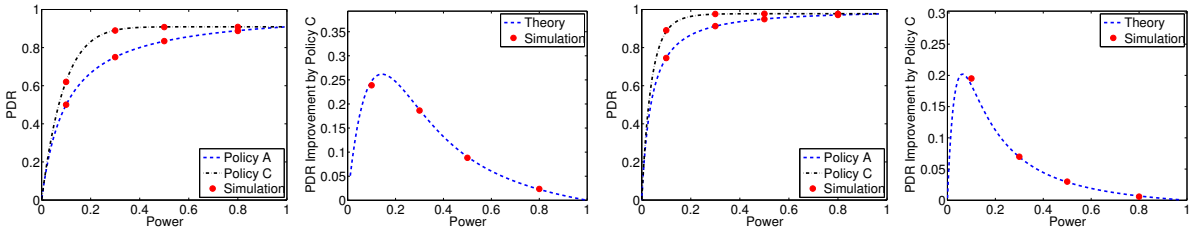
(a) Policy A (b) Policy B ($p_1 = 0, p_2 = 1$) (c) Policy D ($m = 0.25$)

Figure 8.2: *Délai de livraison* dans un réseau de $N = 100$ noeuds où l'on fait varier les caractéristiques de *Mobilité* ($\mu_\lambda = 1$ et $CV_\lambda \in [0, 3]$) pour trois politiques d'égoïsme différentes en utilisant un routage épidémique. La valeur théorique du délai de livraison pour deux paramètres (p_0) pour chaque politique d'égoïsme sont présentés avec des lignes en semi pointillés alors que les moyennes pour la simulation correspondantes sont présentées en pointillés.



(a) Synthetic: uniform policy (b) Synthetic: non - uniform policies (c) Infocom (d) Sigcomm

Figure 8.3: Compromis entre consommation énergétique et délai de livraison d'un message. Simulations synthétiques avec une politique d'égoïsme (a) uniforme et (b) non-uniforme. Simulations sur des trace réelles de (c) Infocom et (d) Sigcomm dans des scénarios de politiques d'égoïsme uniformes et non-uniformes.



(a) $M = 1$, $T = \frac{10}{\mu_\lambda}$ (b) $M = 1$, $T = \frac{10}{\mu_\lambda}$ (c) $M = 5$, $T = \frac{5}{\mu_\lambda}$ (d) $M = 5$, $T = \frac{5}{\mu_\lambda}$

Figure 8.4: (a),(c) Ratio de probabilité de livraison d'un contenu pour la politique d'égoïsme A (bleu) et la politque d'égoïsme C (noir) pour différent niveaux de consommation énergétique. (b),(d) Différence relative du ratio de probabilité de livraisons entre la Politique d'égoïsme C et A, i.e. $\frac{PDR_C - PDR_A}{PDR_A}$. charactéristiques de mobilité: $\mu_\lambda = 1$; (a),(b) $CV_\lambda = 1$ and (c),(d) $CV_\lambda = 2$.

Table 8.2: Valeurs de $c(N, L)$ pour trois protocoles de routage.

Epidemic	$c(N,L) \approx \frac{\ln(N)}{N}$
2-hop	$c(N,L) = \sum_{k=1}^{N-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+2} \cdot (N-k-1)!}$
SnW	$c(N,L) \leq \sum_{k=1}^{L-1} \frac{k^2 \cdot (N-1)!}{(N-1)^{k+2} \cdot (N-k-1)!} + \left(\frac{L}{N-1} + \frac{1}{L}\right) \frac{(N-1)!}{(N-1)^L \cdot (N-L-1)!}$

Table 8.3: Politiques d'égoïsme social.

Politique A	$p(\lambda) = p_0$	
Politique B	$p(\lambda) = \begin{cases} p_1 & : \lambda \leq \lambda_0 \\ p_2 & : \lambda > \lambda_0 \end{cases}$	$\overline{F}_\lambda(\lambda_0) = p_0$
Politique C	$p(\lambda) = \begin{cases} p_1 & : \lambda \leq \lambda_0 \\ p_2 \cdot \frac{\lambda_0}{\lambda} & : \lambda > \lambda_0 \end{cases}$	$\overline{F}_\lambda(\lambda_0) = p_0$
Politique D	$p(\lambda) = p_0 \cdot (1 - e^{-m \cdot \lambda})$	

8.3 Chapitre 4 – Modélisation et analyse de l'hétérogénéité du trafic de la communication dans les RSM.

L'hétérogénéité de la mobilité et son impact dans les RSM ont été largement étudié à la fois par des simulations et des analyses. Toutefois, cela n'a pas été le cas avec l'hétérogénéité du trafic de communication. Dans la grande majorité des travaux antérieurs sur l'évaluation des performances des protocoles de routage, le trafic est supposé homogène, c'est-à-dire que chaque paire de noeuds est tout aussi probable d'être la source et la destination d'un message. Cette hypothèse ne caractérise la plus part des cas où les caractéristiques sociales des noeuds peuvent affecter de manière significative la demande de trafic entre eux. Motivé par cette absence de travaux connexes, dans ce chapitre, nous explorons l'effet de l'hétérogénéité du trafic de communication dans les RSM.

Nous dérivons des résultats qui montrent les effets conjoints du trafic et de la mobilité sur les mécanismes de communication. Parmi les différentes conclusions résultant de notre analyse, nous identifions les conditions dans lesquelles l'hétérogénéité rend la valeur ajoutée de l'utilisation de relais supplémentaires plus ou moins utile. De plus, nous confirmons l'intuition que l'hétérogénéité ferme l'écart de performance entre les différentes protocoles: Il rend le routage plus difficile dans certains cas, et moins nécessaires dans d'autres. Nous croyons que ces premiers résultats d'analyse sur les effets de l'hétérogénéité du trafic constituent une étape importante vers une meilleure conception des protocoles et de l'évaluation de la faisabilité des applications dans les RSM.

8.3.1 Trafic de la communication: Le modèle

En premier lieu, nous avons essayé d'identifier quelles caractéristiques de l'hétérogénéité ont un effet sur la performance. Dans ce sens, dans la figure 8.5 nous montrons à l'aide de simulations que la performance n'est affectée que lorsque le trafic est en corrélation avec la mobilité.

Basé sur ces résultats, nous proposons le modèle suivant capable de décrire une large gamme de modèles de trafic non-uniformes.

Définition 8.3.1 (Trafic de la communication hétérogène)**.** *La demande de trafic entre une paire de noeuds $\{i, j\}$, est une variable aléatoire τ_{ij}, avec $E[\tau_{ij}] = \tau(\lambda_{ij})$, où $\tau(\cdot)$ est une fonction continue: de $\mathbb{R}^+$ à $\mathbb{R}^+$.*

8.3.2 Analyse

Nous montrons la proposition suivante:

Proposition 8.3.1. *La densité de probabilité f_τ du taux de contact entre la source et la destination $\{s, d\}$ d'un message aléatoire, dans un Réseau de Contact Hétérogène (Def. 8.1.2) avec trafic de la communication hétérogène (Def. 8.3.1), converge comme suit:*

$$f_\tau(x) \xrightarrow{p} \frac{1}{\mathcal{C}} \cdot \tau(x) \cdot f_\lambda(x) \tag{8.9}$$

où $f_\tau(x)dx = P\{\lambda_{sd} \in [x, x + dx)\}$, $\xrightarrow{p}$ indique la convergence en probabilité, et $\mathcal{C} = E[\tau(\lambda)] = \int_0^\infty \tau(x) f_\lambda(x) dx$ est une constante de normalisation.

A partir de la proposition 8.3.1, nous dérivons des expressions analytiques pour calculer l'effet conjoint du trafic et de la mobilité dans les performances de la transmission directe ("DT"), et du routage assisté de L relais ("R"). Les résultats génériques et les résultats d'une étude de cas réaliste (avec $f_\lambda(x) \sim \Gamma(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ et $\tau(x) = c \cdot x^k$, $c > 0$) sont donnés à la table 8.4

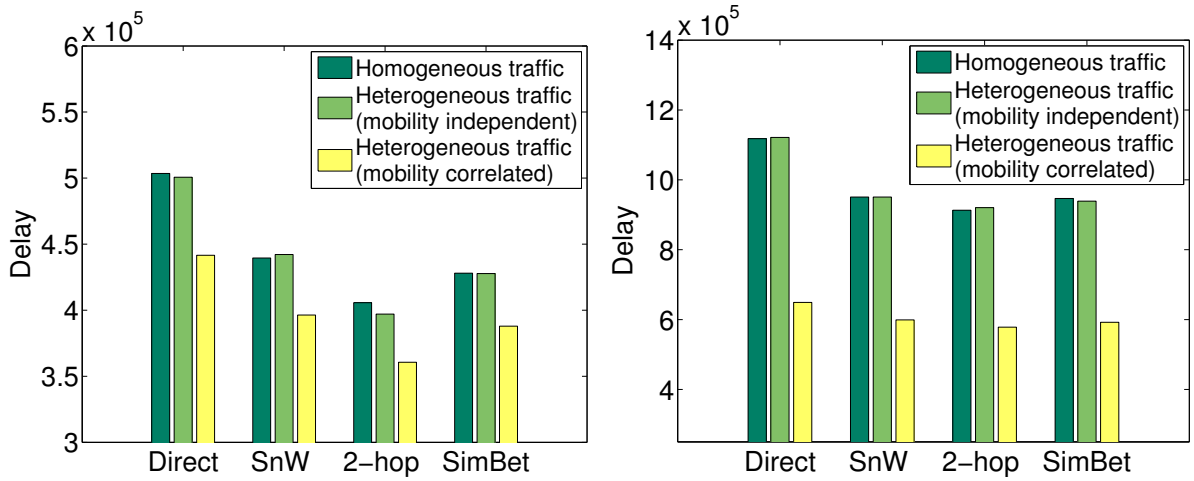
8.3.3 Validation du modèle

Nous validons nos résultats analytiques par des simulations (synthétiques et sur des traces réelles). Nous présentons quelques résultats dans la Figure 8.6 et Figure 8.7, pour le ratio des délais $R = \frac{E[T_R]}{E[T_{DT}]}$ et la probabilité $\mathbf{P_{(src.)}}$ qui est la probabilité qu'un message soit livré à la destination par le noeud source, et non que par l'un des relais.

8.3.4 Publications

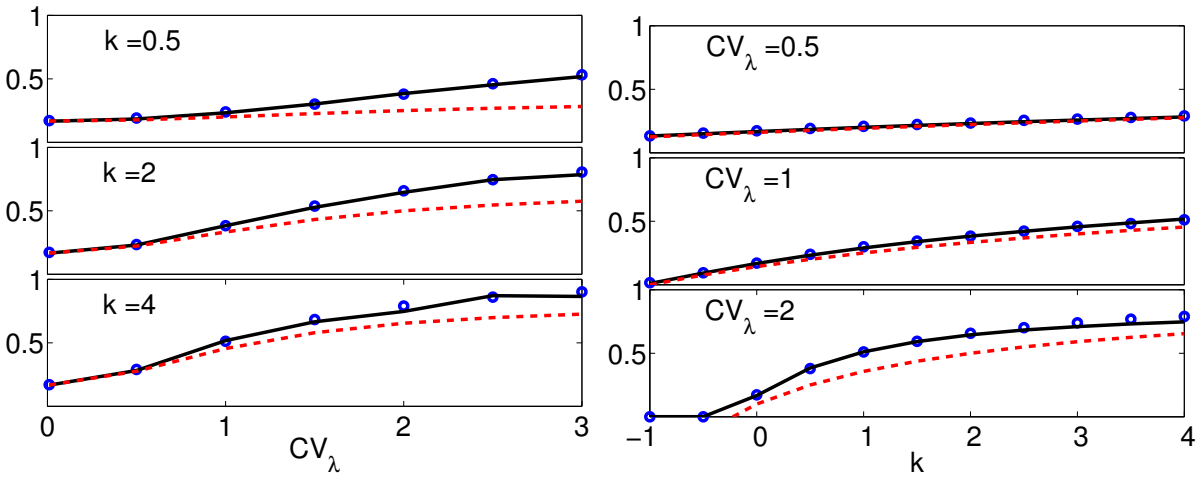
Le travail dans ce chapitre a été publié dans:

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Modelling and analysis of communication traffic heterogeneity in opportunistic networks", IEEE Transactions on Mobile Computing, pending major revision, October 2014.*



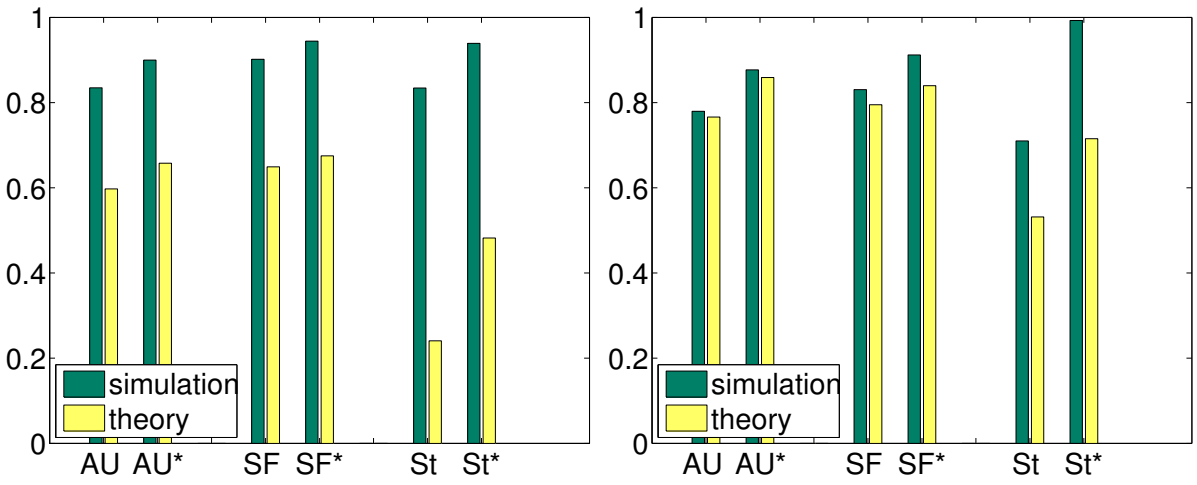
(a) Gowalla dataset (b) Strathclyde dataset

Figure 8.5: Les délais des protocoles de routage:*Direct Transmission*, *Spray and Wait* (*SnW*), *2-hop*, and *SimBet*, sur les traces réelles (a) Gowalla et (b) Strathclyde.



(a) $P_{(src.)}$ vs. CV_λ (b) $P_{(src.)}$ vs. k

Figure 8.6: La probabilité $P_{(src.)}$ dans les scénarios avec variable (a) mobilité et (b) trafic. Les résultats de simulation sont indiqués par des cercles, les prédictions théoriques sont indiqués avec des lignes continues, et les bornes inférieures P_{min} sont indiqués avec des lignes en pointillés.



(a) $R = \frac{E[T_R]}{E[T_{DT}]}$ (b) $P_{(src.)}$

Figure 8.7: Les résultats de simulation pour les R et $P_{(src.)}$, et les prédictions théoriques pour les scénarios du trafic homogène et hétérogène ($*$).

Table 8.4: Les délais et les probabilités de livraison pour le transmission directe ("DT") et le routage assisté de L relais ("R").

Direct Transmission	**Relay-Assisted**
Cas Générique:	
$E[T_{DT}] = \frac{1}{E[\tau(\lambda)]} \cdot E\left[\frac{\tau(\lambda)}{\lambda}\right]$	$E[T_R] = \frac{1}{E[\tau(\lambda)]} \cdot \int_0^\infty \int_0^\infty \frac{\tau(x)}{x+y} \cdot f_\lambda(x)dx \cdot f_R(y)dy$
$P\{T_{DT} \leq t\} = 1 - \frac{E[\tau(\lambda) \cdot e^{-\lambda \cdot t}]}{E[\tau(\lambda)]}$	$P\{T_R \leq t\} = 1 - \frac{E[\tau(\lambda) \cdot e^{-\lambda \cdot t}]}{E[\tau(\lambda)]} \cdot \int_0^\infty e^{-y \cdot t} \cdot f_R(y)dy$
Mobilité $f_\lambda(x) \sim \Gamma(x; \alpha, \beta)$, Trafic $\tau(x) = c \cdot x^k$:	
$E[T_{DT}] = \frac{1}{\mu_\lambda} \cdot \frac{1}{1 + (k-1) \cdot CV_\lambda^2}$	$E[T_R] \geq \frac{1}{\mu_\lambda} \cdot \frac{1}{1 + k \cdot CV_\lambda^2 + L}$
$P\{T_{DT} \leq t\} = 1 - \left(1 + \mu_\lambda \cdot CV_\lambda^2 \cdot t\right)^{-\frac{1+k \cdot CV_\lambda^2}{CV_\lambda^2}}$	$P\{T_R \leq t\} = 1 - \left(1 + \mu_\lambda \cdot CV_\lambda^2 \cdot t\right)^{-\frac{1+k \cdot CV_\lambda^2 + L}{CV_\lambda^2}}$

8.4 Chapitre 5 – Effets des modes de popularité de contenu et de disponibilité de contenu

Dans le Chapitre 4 nous nous sommes concentrés sur les effets de l'hétérogénéité du trafic à la communication de bout-à-bout. Néanmoins, dans de nombreuses applications le trafic n'est pas entre une paire de noeuds (i.e. source-destination), mais l'objectif principal est de distribuer le contenu aux utilisateurs intéressés. Dans ce cas, l'hétérogénéité figure parmi les groupes de noeuds participant à la distribution de différentes contenus. Plus précisément, les modèles d'intérêt, c'est-à-dire le nombre de noeuds qui sont intéressés à chaque contenu (*popularité*), ainsi que le nombre d'utilisateurs qui peuvent fournir un contenu (*disponibilité*), affectent les performances des applications. Ainsi, dans ce chapitre, nous établissons un cadre analytique pour étudier les effets de ces facteurs sur le délai et la probabilité de livraison dans le RSM.

8.4.1 Le modèle du trafic

D'abord, nous proposons un modèle analytique simple qui se applique à une gamme de motifs de popularité de contenu vu dans les réseaux réels; à notre connaissance, ce est le premier effort indépendante de l'application dans cette direction.

8.4.1.1 Popularité de contenu

Nous supposons un réseau avec N noeuds, et on note l'ensemble des noeuds que $\mathcal{N}$. On note le cas où un noeud $i \in \mathcal{N}$ est *intéressé* par un contenu $\mathcal{M}$ (ou, de façon équivalente, i demande $\mathcal{M}$), comme:

$$i \to \mathcal{M}$$

On note l'ensemble de tous les contenus, dans lequel les noeuds sont intéressés, comme:

$$\mathbf{M} = \{\mathcal{M} : \exists i \in \mathcal{N}, i \to \mathcal{M}\}. \quad |\mathbf{M}| = M$$

où $|\cdot|$ indique la cardinalité d'un ensemble.

Définition 8.4.1 (Popularité de contenu)**.** *Nous définissons la popularité d'un contenu $\mathcal{M}$ comme le nombre de noeuds $N_p^{(\mathcal{M})}$ qui sont intéressés à lui:*

$$N_p^{(\mathcal{M})} = |\mathcal{C}_p^{(\mathcal{M})}|, \text{ où } \mathcal{C}_p^{(\mathcal{M})} = \{i \in \mathcal{N} : i \to \mathcal{M}\} \tag{8.10}$$

Plus, on note le pourcentage de contenu avec une valeur de popularité n comme

$$P_p(n) = \frac{1}{M} \sum_{\mathcal{M} \in \boldsymbol{M}} \mathcal{I}_{N_p^{(\mathcal{M})}=n}, \quad n \in [0, N] \tag{8.11}$$

où $\mathcal{I}_{N_p^{(\mathcal{M})}=n} = 1$ quand $N_p^{(\mathcal{M})} = n$ et 0 dans un autre cas.

8.4.1.2 Disponibilité de contenu

Nous supposons que d'une demande de contenu est terminée, quand un noeud qui détient le contenu demandé est rencontré directement. On note le cas où un noeud i qui détient (une copie du) le contenu $\mathcal{M}$ comme

$$i \leftarrow \mathcal{M}$$

et nous définissons la disponibilité d'un contenu $\mathcal{M}$ comme

Définition 8.4.2 (Disponibilité de contenu)**.** *Nous définissons la disponibilité d'un contenu $\mathcal{M}$ comme le nombre de noeuds $N_a^{(\mathcal{M})}$ qui détient le contenu:*

$$N_a^{(\mathcal{M})} = |\mathcal{C}_a^{(\mathcal{M})}|, \text{ où } \mathcal{C}_a^{(\mathcal{M})} = \{i \in \mathcal{N} : i \leftarrow \mathcal{M}\} \tag{8.12}$$

8.4.2 Analyse

Sur la base de notre modèle, et les hypothèses suivantes

Hypothèse 8.4.1. *La popularité $N_p^{(\mathcal{M})}$ et la disponibilité $N_a^{(\mathcal{M})}$ d'un contenu $\mathcal{M}$ ne changent pas au cours du temps.*

Hypothèse 8.4.2. *L'ensembles des demandeurs $\mathcal{C}_p^{(\mathcal{M})}$ et détenteurs $\mathcal{C}_a^{(\mathcal{M})}$ d'un contenu $\mathcal{M}$ sont indépendants de la mobilité.*

nous pouvons prouver que

Lemme 8.4.1. *La probabilité que une demande aléatoire pour un contenu de popularité égale à n est donnée par*

$$P_p^{req.}(n) = \frac{n}{E_p[n]} \cdot P_p(n)$$

où $E_p[n] = \sum_n n \cdot P_p(n)$ est la popularité moyenne.

Lemme 8.4.2. *La probabilité que une demande aléatoire pour un contenu de disponibilité égale à m est donnée par*

$$P_a^{req.}(m) = \frac{E_p[n \cdot g(m|n)]}{E_p[n]}$$

Ensuite, en utilisant les lemmes ci-dessus, nous dérivons des expressions de la performance d'un mécanisme de distribution de contenu:

Résultat 8.4.1. *Le délai moyen d'accès au contenu peut être calculée avec l'expression*

$$E[T_{\mathcal{M}}] = \frac{1}{E_p[n]} \cdot E_p\left[n \cdot \sum_m E_{m\lambda}\left[\frac{1}{x}\right] \cdot g(m|n)\right]$$

Résultat 8.4.2. *La probabilité d'un contenu d'être accessible avant un temps TTL peut être calculée avec l'expression*

$$P\{T_{\mathcal{M}} \leq TTL\} = 1 - \frac{E_p\left[n \cdot \sum_m E_{m\lambda}\left[e^{-x \cdot TTL}\right] g(m|n)\right]}{E_p[n]}$$

Nous derivons aussi des bornes pour la prédiction de la performance qui nécessitent peu de connaissances sur les caractéristiques du réseau et des motifs d'intérêt, et ils peuvent donc être utilisés en situations réelles, pour la conception de protocoles, l'optimisation en ligne, etc.

Théorème 8.4.1.

$$E[T_{\mathcal{M}}] \geq \frac{1}{\mu_\lambda \cdot E_p[n]} \cdot E_p\left[\frac{n}{\overline{g}(n)}\right]$$

Théorème 8.4.2.

$$P\{T_{\mathcal{M}} \leq TTL\} \leq 1 - \frac{1}{E_p[n]} \cdot E_p\left[n \cdot e^{-\overline{g}(n)\cdot\mu_\lambda \cdot TTL}\right]$$

La validation de nos résultats par des simulations, montre une précision importante (voir, par exemple, la Figure 8.8)

8.4.3 Une étude de cas: déchargement de données mobiles

Nous appliquons notre cadre au problème de déchargement de données mobile et nous fournissons quelques idées initiales pour l'optimisation de sa performance.

Par exemple, on montre que l' allocation optimale qui minimise le délai est donnée par

Résultat 8.4.3. *Le délai minimum, sous la contrainte d'un nombre moyen de $c_{\mathcal{M}}$ copies par contenu, i.e.*

$$\min\{E[T_{\mathcal{M}}]\} \quad s.t. \quad \sum_{\mathcal{M}} N_a^{(\mathcal{M})} = M \cdot c_{\mathcal{M}} \ , \ N_a^{(\mathcal{M})} \geq 0$$

il peut être atteint lorsque la fonction d'allocation, $g(m|n)$, est déterministe et égale à

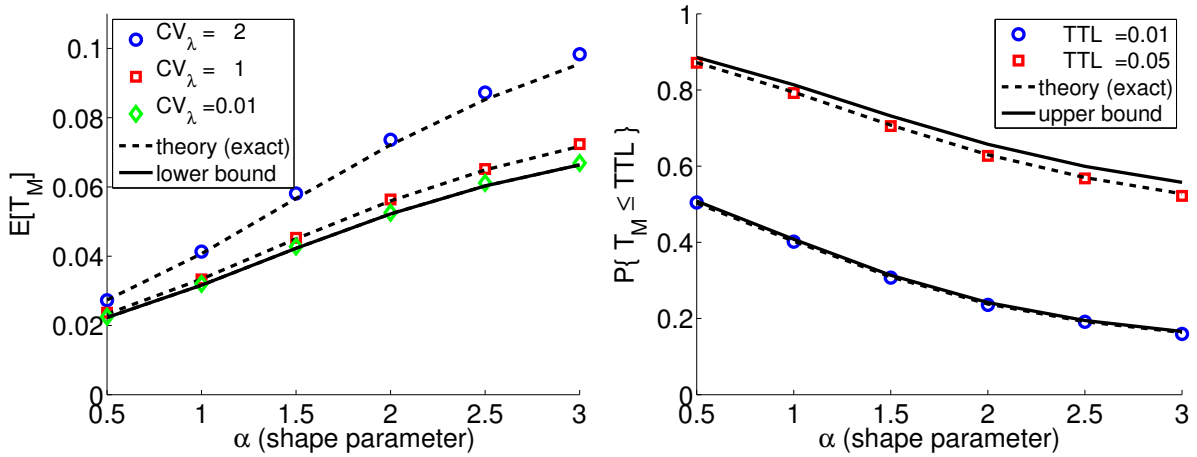
$$\rho^*(n) = \frac{c_{\mathcal{M}}}{E_p[\sqrt{n}]} \cdot \sqrt{n}$$

Ce résultat est vérifié ainsi dans les simulations, où nous avons utilisé des traces de mobilité réelle; nous exposons le résultat dans la Figure 8.9.

8.4.4 Publications

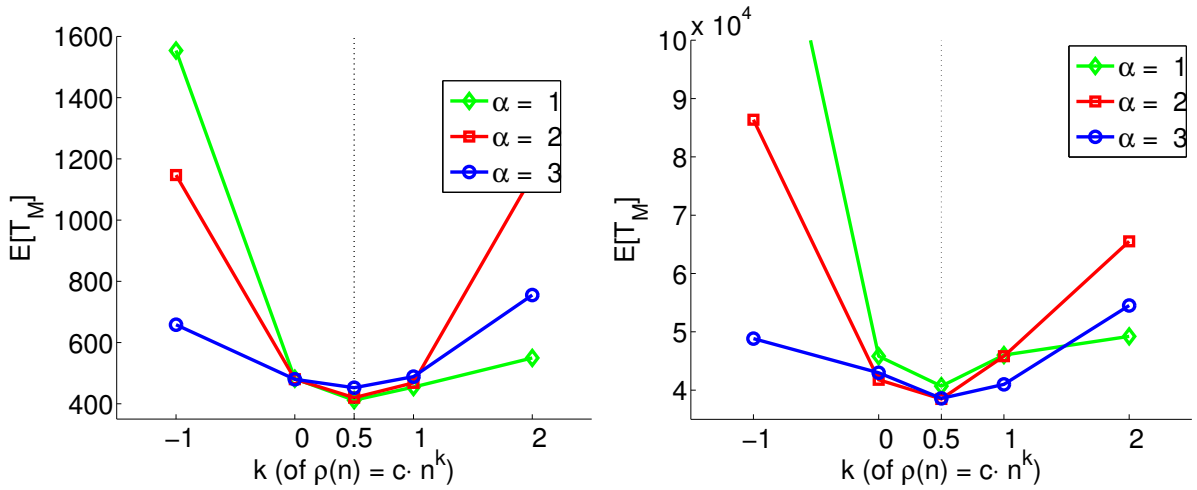
Le travail dans ce chapitre a été publié dans:

- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Not all content is created equal: Effect of popularity and availability for content-centric opportunistic networking", Proc. 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'14), August 11-14, 2014, Philadelphia, PA, USA.*
- *Pavlos Sermpezis, Thrasyvoulos Spyropoulos, "Effects of content popularity in the performance of content-centric opportunistic networking: An analytical approach and applications", IEEE/ACM Transactions on Networking, submitted, September 2014.*



(a) $E[T_{\mathcal{M}}]$ (b) $P\{T_{\mathcal{M}} \leq TTL\}$

Figure 8.8: (a) $E[T_{\mathcal{M}}]$ et (b) $P\{T_{\mathcal{M}} \leq TTL\}$ dans les scénarios avec la popularité de contenu variable (α: paramètre de forme) et $\rho(n) = 0.2 \cdot n$.



(a) TVCM (b) Cabspotting

Figure 8.9: Le délai $E[T_{\mathcal{M}}]$ des différentes politiques d'allocation $\rho(n) = c_k \cdot n^k$, où $c_k = \frac{c_{\mathcal{M}}}{E_p[n^k]}$.

8.5 Chapitre 6 – "Offloading on the Edge": Analyse et optimisation du stockage local des données et de déchargement dans HetNets

Basé sur l'analyse du Chapitre 5 pour les effets de la hétérogénéité du trafic, dans ce chapitre, on se concentre sur une application de la distribution de contenu, *déchargement de données mobiles*, qui a récemment attiré beaucoup d'attention, en raison de l'augmentation rapide de la demande de trafic de données qui a surchargé les réseaux cellulaires.

8.5.1 "Offloading on the Edge ": Le modèle

On propose un modèle analytique pour explorer comment le stockage local des données et la communication opportuniste par "edge" noeuds (c' est à dire les noeuds mobiles et les "small-cells") pourraient aider à décharger du trafic dans un réseau hétérogène (HetNet).

D' abord, on définisse les ensembles de "edge" noeuds qui sont associés dans le processus de déchargement:

Définition 8.5.1. *Un demandeur d'un contenu est un noeud mobile (MN) qui (a) est intéressé par le contenu et (b) ne l'a pas encore reçu. On note l'ensemble des demandeurs au temps t comme* $\mathcal{R}(t)$.

Définition 8.5.2. *Un détenteur d'un contenu est un "edge" noeud (SC ou MN) qui stocke le contenu et le transmettra à ses demandeurs. On note l'ensemble des détenteurs au temps t comme* $\mathcal{H}(t)$.

Les coûts impliqués dans chaque phase de la mécanisme "offloading on the edge" sont:

– ***Coûts de placement initial:*** C_{BH}, C_{BS}.

- Un contenu est placé à un SC par un transmission de backhaul (filaire ou sans fil), et on note ce coût par placement comme C_{BH}.
- Un placement de contenu à un MNs a lieu à une transmission de BS (macro-cell). On note ce coût de transmission C_{BS}.

– ***Les coûts de déchargement opportunistes:*** C_{SC}, C_{D2D}.

- Le coût d'une transmission SC-MN: C_{SC}.
- Le coût d'une transmission MN-MN (ou D2D): C_{D2D}.

– ***Le coût de la livraison retardée:*** $C_{BS}^{(TTL)}$.

- Le coût d'une transmission de distribution retardé est $C_{BS}^{(TTL)}$.

8.5.2 Analyse

En utilisant une analyse basée sur de approximations "Mean-field" et "Fluid-model", on peut d'abord montrer que

Lemme 8.5.1. *Le "fluid-limit" approximation déterministe pour le nombre attendu des détenteurs ($H(t)$) et les demandeurs ($R(t)$) au moment de t, est*

$$H(t)=H_0 \cdot \frac{(p_c \cdot R_0 + H_0) \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot t}}{p_c \cdot R_0 + H_0 \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot t}}$$
$$R(t)=R_0 \cdot \frac{p_c \cdot R_0 + H_0}{p_c \cdot R_0 + H_0 \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot t}}$$

où $H_0 = H(0^+)$ et $R_0 = R(0^+)$.

Puis, basé sur le lemme 8.5.1, on peut calculer la performance du mécanisme "offloading on the edge"

Résultat 8.5.1 (**Probabilité de livraison**). *La probabilité pour un contenu à être livré à un demandeur par le temps t est donnée par*

$$P\{T_d \leq t\} = 1 - \frac{p_c \cdot R_0 + H_0}{p_c \cdot R_0 + H_0 \cdot e^{\mu_\lambda \cdot (p_c \cdot R_0 + H_0) \cdot t}}$$

où $H_0 = H(0^+)$ et $R_0 = R(0^+)$.

On peut également calculer le coût de déchargement d'un seul contenu

Résultat 8.5.2. *Le coût de déchargement d'un seul contenu par le mecanisme "offloading on the edge", est donnée par*

$$\begin{aligned} C =& C_{BH} \cdot H_{SC}(0) + C_{BS} \cdot H_{MN}(0) \\ &+ (C_{SC} \cdot q + C_{D2D} \cdot (1-q)) \cdot R_0 \cdot P\{T_d \leq TTL\} \\ &+ C_{BS}^{(TTL)} \cdot R_0 \cdot (1 - P\{T_d \leq TTL\}) \end{aligned}$$

où $q = \frac{H_{SC}(0) \cdot \ln\left(\frac{H(TTL)}{H_0}\right)}{p_c \cdot (R_0 - R(TTL))}$, et $P\{T_d \leq TTL\}$, $H(TTL)$ et $R(TTL)$ sont donnée par le Lemme 8.5.1 et le Résultat 8.5.1.

8.5.3 Applications: Optimisation du coût

On suppose que le fournisseur de contenu doit livrer $M \geq 1$ contenus à leurs demandeurs. On note l'ensemble des contenus comme $\mathcal{M}$ ($M = |\mathcal{M}|$). Si on note le coût de la livraison d'un contenu $\theta \in \mathcal{M}$ (qui est donnée par le Résultat 8.5.2)) comme C^θ, on peut exprimer le problème d'optimisation de coût *totale* comme

Problème 8.5.1.

$$\begin{aligned} \min_{\overline{H}_{SC},\, \overline{H}_{MN},\, \overline{TTL}} \ & \left\{\textstyle\sum_{\theta \in \mathcal{M}} C^\theta\right\} \\ s.t. \ \ & \forall \theta \in \mathcal{M}: \ 0 \leq H_{SC}^\theta(0) \leq N_{SC} \\ & 0 \leq H_{MN}^\theta(0) \leq R^\theta(0) \\ & T_{min} \leq TTL^\theta \leq T_{max} \\ & \textit{and} \qquad \sum_{\theta \in \mathcal{M}} H_{SC}^\theta(0) \leq \sum_{i \in \mathcal{SC}} Q(i) \end{aligned}$$

où $\overline{H_{SC}}$, $\overline{H_{MN}}$ *et* $\overline{TTL}$ *ils dénotent les vecteurs ayant des composantes* $H^{\theta}_{SC}(0)$, $H^{\theta}_{MN}(0)$ *et* TTL^{θ} $(\theta \in \mathcal{M})$*; et* $Q(i)$ *est la capacité de stockage (en nombre de contenu) d'un noeud "small-cell"* i*.*

En utilisant des méthodes bien connues, nous pouvons résoudre le problème d'optimisation ci-dessus. Notamment, dans certains cas, on peut également calculer des expressions analytiques et forme fermée:

8.5.3.1 Déchargement par SCs

Résultat 8.5.3. *Dans un scénario de "Déchargement par SCs" (*$p_c = 0$, $H_{MN}(0) = 0$*), l'allocation initiale* $\overline{H_{SC}}$ *qui minimise le coût total, est donnée par*

$$H^{\theta}_{SC}(0) = \begin{cases} N_{SC} & , R^{\theta}(0) > U \\ \frac{1}{\gamma} \cdot \ln\left(\frac{1}{L} \cdot R^{\theta}(0)\right) & , L \leq R^{\theta}(0) \leq U \\ 0 & , R^{\theta}(0) < L \end{cases}$$

où $\gamma = \mu_{\lambda} \cdot TTL$, $L = \frac{1}{\gamma \cdot \Phi} \cdot \left(1 + \frac{\lambda_0}{C_{BH}}\right)$, $U = L \cdot e^{\gamma \cdot N_{SC}}$, $\Phi = \frac{C^{(TTL)}_{BS} - C_{SC}}{C_{BH}}$, *et*

$$\lambda_0 = \inf\left\{\lambda_0 \geq 0 : \sum_{\theta \in \mathcal{M}} H^{\theta}_{SC}(0) \leq \sum_{i \in \mathcal{SC}} Q(i)\right\}$$

8.5.3.2 Déchargement par MNs

Résultat 8.5.4. *Dans un scénario de "Déchargement par MNs" (*$p_c > 0$, $H_{SC}(0) = 0$*), l'allocation initiale* $\overline{H_{MN}}$ *qui minimise le coût total, est donnée par*

$$H^{\theta}_{MN}(0) = \begin{cases} R^{\theta}(0) & , R^{\theta}(0) \leq OPT^{\theta} \\ OPT^{\theta} & , 0 \leq OPT^{\theta} < R^{\theta}(0) \\ 0 & , OPT^{\theta} < 0 \end{cases}$$

où

$$OPT^{\theta} = \frac{R^{\theta}(0) \cdot \left(\sqrt{\Phi'} \cdot e^{\frac{1}{2}\gamma \cdot p_c \cdot R^{\theta}(0)} - 1\right)}{e^{\gamma \cdot p_c \cdot R^{\theta}(0)} - 1}$$

et $\Phi' = \frac{C^{(TTL)}_{BS} - C_{D2D}}{C_{BS} - C_{D2D}}$ *et* $\gamma = \mu_{\lambda} \cdot TTL$*.*

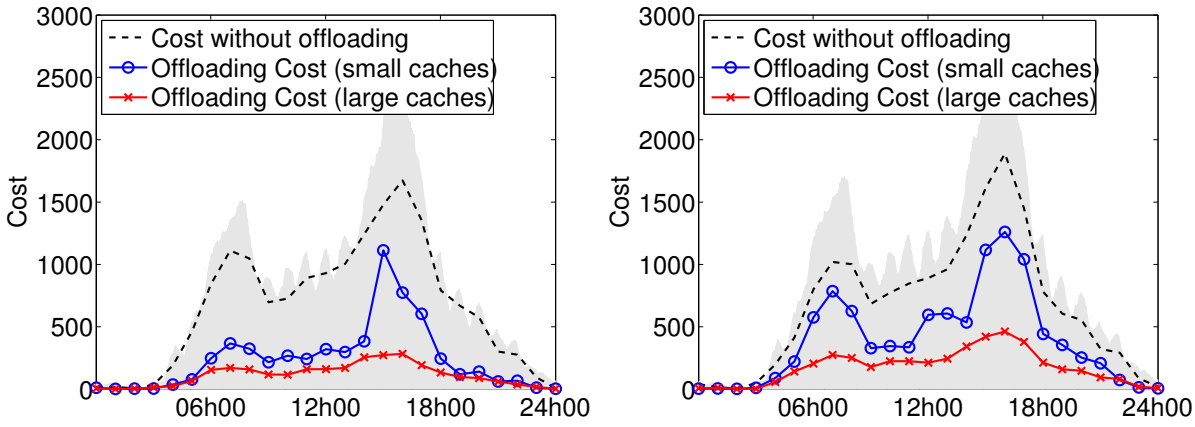
8.5.4 Les résultats des simulations

On évalue l'efficacité de coût de "offloading on the edge" par des simulations à scénarios avec des motifs de la demande de trafic réalistes. Nous présentons les résultats à Fig. 8.10 et Fig. 8.11.

8.5.5 Publications

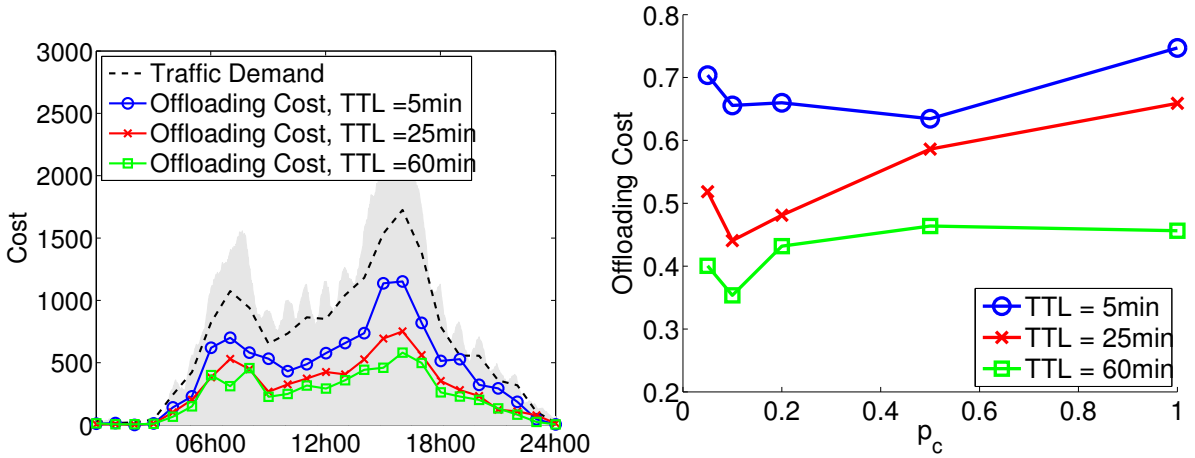
Le travail dans ce chapitre a été publié dans:

- *Pavlos Sermpezis, Luigi Vigneri, Thrasyvoulos Spyropoulos, "Offloading on the Edge: Analysis and optimization of local data storage and offloading in HetNets", Research Report RR-14-297, Eurecom, December 2014.*



(a) $TTL = 25min$ (b) $TTL = 5min$

Figure 8.10: La demande de trafic et le coût de déchargement (par SCs) sur une période de $24h$.



(a) Offloading Cost vs. Time (b) Offloading Cost vs. p_c

Figure 8.11: (a) La demande de trafic et le coût de déchargement (par MNs, avec $p_c = 0.1$) sur une période de $24h$. (b) Le coût de déchargement total sur une période de $24h$, normalisée au coût total sans déchargement.

Bibliography

[1] Akamai, "Swisscom and akamai enter into a strategic partnership," Press Release, March 2013.

[2] D. Aldous and J. Fill, "Reversible markov chains and random walks on graphs. (monograph in preparation.)," *http://stat-www.berkeley.edu/users/aldous/RWG/book.html*.

[3] J. Andrews, "Seven ways that hetnets are a cellular paradigm shift," *Communications Magazine, IEEE*, vol. 51, no. 3, pp. 136–144, March 2013.

[4] A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. Springer, 2007.

[5] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–1, 2014.

[6] A. Balasubramanian, B. Levine, and A. Venkataramani, "Dtn routing as a resource allocation problem," in *Proc. ACM SIGCOMM*, 2007.

[7] M. Barthelemy, A. Barrat, R. Pastor-Satorras, and A. Vespignani, "Dynamical patterns of epidemic outbreaks in complex heterogeneous networks," *Journal of Theor. Biology*, vol. 235, no. 2, pp. 275–288, 2005.

[8] S. Batabyal and P. Bhaumik, "Estimators for global information in mobile opportunistic network," in *Proc. IEEE ANTS*, 2013.

[9] C. Boldrini, M. Conti, and A. Passarella, "Performance modelling of opportunistic forwarding with imprecise knowledge," in *Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt), 2012 10th International Symposium on*, May 2012, pp. 216–223.

[10] ——, "Design and performance evaluation of contentplace, a social-aware data dissemination system for opportunistic networks," *Computer Networks*, vol. 54, no. 4, pp. 589–604, 2010.

[11] ——, "Modelling social-aware forwarding in opportunistic networks," in *Proc. IFIP PERFORM*. Springer-Verlag, 2011.

[12] ——, "Less is more: Long paths do not help the convergence of social-oblivious forwarding in opportunistic networks," in *Proc. ACM MobiOpp*, 2012.

[13] C. Boldrini and A. Passarella, "Hcmm: Modelling spatial and temporal properties of human mobility driven by users' social relationships," *Computer Communications*, vol. 33, no. 9, pp. 1056–1074, 2010.

[14] V. Borrel, F. Legendre, M. Dias de Amorim, and S. Fdida, "Simps: Using sociology for personal mobility," *Networking, IEEE/ACM Transactions on*, vol. 17, no. 3, pp. 831–842, June 2009.

[15] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and zipf-like distributions: evidence and implications," in *Proc. IEEE INFOCOM*, 1999.

[16] H. Cai and D. Y. Eun, "Crossing over the bounded domain: from exponential to power-law intermeeting time in mobile ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 17, no. 5, pp. 1578–1591, Oct. 2009.

[17] H. Cai, I. Koprulu, and N. Shroff, "Exploiting double opportunities for deadline based content propagation in wireless networks," in *Proc. IEEE INFOCOM*, 2013.

[18] A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott, "Impact of human mobility on the design of opportunistic forwarding algorithms," in *Proc. IEEE INFOCOM*, 2006.

[19] ——, "Impact of human mobility on opportunistic forwarding algorithms," *IEEE Trans. on Mobile Computing*, vol. 6, no. 6, 2007.

[20] A. Chaintreau, J.-Y. Le Boudec, and N. Ristanovic, "The age of gossip: spatial mean field regime," in *SIGMETRICS 2009*, pp. 109–120.

[21] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: a survey," *Communications Magazine, IEEE*, vol. 46, no. 9, pp. 59–67, September 2008.

[22] B. B. Chen and M. C. Chan, "Mobicent: a credit-based incentive system for disruption tolerant network," in *Proc. IEEE INFOCOM*, 2010.

[23] CISCO, "Cisco visual networking index: Global mobile data traffic forecast update, 2013–2018," Tech. Rep., 2014.

[24] E. Cohen and S. Shenker, "Replication strategies in unstructured peer-to-peer networks," in *Proc. ACM SIGCOMM*, 2002.

[25] V. Conan, J. Leguay, and T. Friedman, "Characterizing pairwise inter-contact patterns in delay tolerant networks," in *Proc. ACM Autonomics*, 2007.

[26] M. Conti, S. Giordano, M. May, and A. Passarella, "From opportunistic networks to opportunistic computing," *Communications Magazine, IEEE*, vol. 48, no. 9, pp. 126–139, sept. 2010.

[27] E. Cooch and G. White, *Program MARK A gentle Introduction*, 11st ed., 2012.

[28] P. Costa, C. Mascolo, M. Musolesi, and G. Picco, "Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks," *IEEE JSAC*, vol. 26, no. 5, 2008.

[29] E. M. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *Proc. ACM MobiHoc*, 2007.

[30] E. Daly and M. Haahr, "Social network analysis for information flow in disconnected delay-tolerant manets," *IEEE Trans. on Mobile Computing*, vol. 8, no. 5, pp. 606–621, May 2009.

[31] N. Eagle and A. Pentland, "Inferring social network structure using mobile phone data," in *Proceedings of the National Academy of Sciences (PNAS)*, vol. Vol 106 (36), 2009, pp. 15 274–15 278.

[32] J. Erman, A. Gerber, K. K. Ramadrishnan, S. Sen, and O. Spatscheck, "Over the top video: The gorilla in cellular networks," in *Proc. ACM IMC*, 2011.

[33] K. Fall, "A delay-tolerant network architecture for challenged internets," in *Proc. ACM SIGCOMM*, 2003.

[34] W. Gao and G. Cao, "User-centric data dissemination in disruption tolerant networks," in *Proc. IEEE INFOCOM*, 2011.

[35] W. Gao, G. Cao, T. La Porta, and J. Han, "On exploiting transient social contact patterns for data forwarding in delay-tolerant networks," *IEEE Trans. on Mob. Computing*, vol. 12, no. 1, pp. 151–165, jan. 2013.

[36] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: a social network perspective," in *Proc. ACM MobiHoc*, 2009.

[37] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proc. CHI*, 2009.

[38] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: A view from the edge," in *Proc. ACM IMC*, 2007.

[39] N. Golrezaei, A. Molisch, A. Dimakis, and G. Caire, "Femtocaching and device-to-device collaboration: A new architecture for wireless video distribution," *IEEE Comm. Magazine*, vol. 51, no. 4, April 2013.

[40] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products, Seventh Edition*, 7th ed. Academic Press, 2007.

[41] M. S. Granovetter, "The Strength of Weak Ties," *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.

[42] S. Grasic and A. Lindgren, "Revisiting a remote village scenario and its DTN routing objective," *Computer Communications*, vol. 48, pp. 133–140, 2014, opportunistic networks.

[43] R. Groenevelt, P. Nain, and G. Koole, "The message delay in mobile ad hoc networks," *Performance Evaluation*, vol. 62, pp. 210–228, 2005.

[44] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad-hoc wireless networks," in *Proc. IEEE INFOCOM*, 2001.

[45] A. Guerrieri, I. Carreras, F. D. Pellegrini, D. Miorandi, and A. Montresor, "Distributed estimation of global parameters in delay-tolerant networks," *Computer Communications*, vol. 33, no. 13, pp. 1472–1482, 2010.

[46] S. Guo, M. Derakhshani, M. Falaki, U. Ismail, R. Luk, E. Oliver, S. U. Rahman, A. Seth, M. Zaharia, and S. Keshav, "Design and implementation of the kiosknet system," *Computer Networks*, vol. 55, no. 1, pp. 264–281, 2011.

[47] A. Gut, *An intermediate course in probability*, 2nd ed. Springer Verlag, 2009.

[48] S. Ha, S. Sen, C. Joe-Wong, Y. Im, and M. Chiang, "Tube: Time-dependent pricing for mobile data," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 42, no. 4, pp. 247–258, 2012.

[49] Z. J. Haas and T. Small, "A new networking model for biological applications of ad hoc sensor networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 14, pp. 27–40, 2006.

[50] B. Han, P. Hui, V. Kumar, M. Marathe, J. Shao, and A. Srinivasan, "Mobile data offloading through opportunistic communications and social participation," *IEEE Trans. on Mob. Comp.,*, vol. 11, no. 5, 2012.

[51] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," *Computer Networks*, vol. 52, no. 14, pp. 2690–2712, 2008.

[52] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing," in *Proc. IEEE INFOCOM*, 2010.

[53] T. Hossmann, P. Carta, D. Schatzmann, F. Legendre, P. Gunningberg, and C. Rohner, "Twitter in disaster mode: Security architecture," in *ACM CoNEXT 2011 - SWID Workshop*.

[54] T. Hossmann, G. Nomikos, T. Spyropoulos, and F. Legendre, "Collection and analysis of multi-dimensional network data for opportunistic networking research," *Comput. Communications*, vol. 35, no. 13, 2012.

[55] T. Hossmann, T. Spyropoulos, and F. Legendre, "Putting contacts into context: mobility modeling beyond inter-contact times," in *Proc. ACM MobiHoc*, 2011.

[56] W. Hsu and A. Helmy, "On nodal encounter patterns in wireless lan traces," *IEEE Transactions on Mobile Computing*, vol. 9, pp. 1563–1577, 2010.

[57] W.-J. Hsu, T. Spyropoulos, K. Psounis, and A. Helmy, "Modeling spatial and temporal dependencies of user mobility in wireless mobile networks," *IEEE/ACM Trans. on Networking*, vol. 17, no. 5, 2009.

[58] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot, "Pocket switched networks and human mobility in conference environments," in *Proc. ACM WDTN*, 2005.

[59] P. Hui and J. Crowcroft, "How small labels create big improvements," in *IEEE PerCom Workshops*, 2007.

[60] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay-tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 10, pp. 1576–1589, 2011.

[61] P. Hui, K. Xu, V. Li, J. Crowcroft, V. Latora, and P. Lio, "Selfishness, altruism and message spreading in mobile social networks," in *Proc. IEEE INFOCOM Workshops*, 2009.

[62] E. Hyytia, J. Virtamo, P. Lassila, J. Kangasharju, and J. Ott, "When does content float? characterizing availability of anchored information in opportunistic content sharing," in *Proc. IEEE INFOCOM*, 2011.

[63] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and scalable distribution of content updates over a mobile social network," in *Proc. IEEE INFOCOM*, 2009.

[64] S. Ioannidis and A. Chaintreau, "On the Strength of Weak Ties in Mobile Social Networks," in *Proc. ACM SNS Workshop*, 2009.

[65] M. Ji, G. Caire, and A. F. Molisch, "Fundamental limits of caching in wireless d2d networks," arXiv preprint, http://arxiv.org/abs/1405.5336, 2014.

[66] K. Johansson, "Cost effective deployment strategies for heterogenous wireless networks," *Doctoral Thesis*, 2007.

[67] P. Juang, H. Oki, Y. Wang, M. Martonosi, L. S. Peh, and D. Rubenstein, "Energy-efficient computing for wildlife tracking: Design tradeoffs and early experiences with zebranet," in *Proc. ACM ASPLOS-X*, 2002.

[68] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnović, "Power law and exponential decay of inter contact times between mobile devices," in *Proc. ACM MobiCom*, 2007.

[69] M. Karaliopoulos, "Assessing the vulnerability of DTN data relaying schemes to node selfishness," *IEEE Communications Letters*, vol. 13, no. 12, pp. 923–925, 2009.

[70] A. Karr, *Probability*, ser. Springer texts in statistics. Springer, 1993.

[71] M. J. Keeling, "The effects of local spatial structure on epidemiological invasions." *Proc. R. Soc. B*, vol. 266, no. 1421, pp. 859–867, 1999.

[72] A. Khelil, C. Becker, J. Tian, and K. Rothermel, "An epidemic model for information diffusion in manets," in *Proc. of ACM MSWiM*, 2002.

[73] Y. Kim, K. Lee, N. B. Shroff, and I. Rhee, "Providing probabilistic guarantees on the time of information spread in opportunistic networks," in *Proc. IEEE INFOCOM*, 2013.

[74] A. Krifa, C. Barakat, and T. Spyropoulos, "Mobitrade: trading content in disruption tolerant networks," in *Proc. ACM CHANTS*, 2011.

[75] J.-Y. Le Boudec, "Modelling the Immune System Toolbox: Stochastic Reaction Models," http://infoscience.epfl.ch/record/98734/files/toolbox.pdf, 2006.

[76] C.-H. Lee and D. Y. Eun, "On the forwarding performance under heterogeneous contact dynamics in mobile opportunistic networks," *IEEE Transactions on Mobile Computing*, vol. 12, no. 6, 2013.

[77] K. Lee, S. Hong, S. J. Kim, I. Rhee, and S. Chong, "Slaw: A new mobility model for human walks," in *IEEE INFOCOM*, 2009.

[78] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: How much can wifi deliver?" in *Proc. ACM CoNEXT*, 2010.

[79] V. Lenders, G. Karlsson, and M. May, "Wireless ad hoc podcasting," in *Proc. IEEE SECON*, 2007.

[80] J. Leskovec, M. Mcglohon, C. Faloutsos, N. Glance, and M. Hurst, "Cascading behavior in large blog graphs," in *In SDM*, 2007.

[81] M. Li, M.-Y. Wu, Y. Li, J. Cao, L. Huang, Q. Deng, X. Lin, C. Jiang, W. Tong, Y. Gui, A. Zhou, X. Wu, and S. Jiang, "Shanghaigrid: an information service grid," *Concurrency and Computation: Practice and Experience*, vol. 18, no. 1, pp. 111–135, 2006.

[82] Q. Li, S. Zhu, and G. Cao, "Routing in socially selfish delay tolerant networks," in *Proc. IEEE INFOCOM*, 2010.

[83] Y. Li, M. Steiner, L. Wang, Z.-L. Zhang, and J. Bao, "Exploring venue popularity in foursquare," in *Proc. IEEE INFOCOM Workshops (NetSciCom)*, 2013.

[84] Y. Li, P. Hui, D. Jin, L. Su, and L. Zeng, "Evaluating the impact of social selfishness on the epidemic routing in delay tolerant networks," *Comm. Letters.*, vol. 14, no. 11, Nov. 2010.

[85] Y. Li, M. Qian, D. Jin, P. Hui, Z. Wang, and S. Chen, "Multiple mobile data offloading through disruption tolerant networks," *IEEE Transactions on Mobile Computing*, vol. 13, no. 7, pp. 1579–1596, 2014.

[86] Y. Li, G. Su, D. Wu, D. Jin, L. Su, and L. Zeng, "The impact of node selfishness on multicasting in delay tolerant networks," *IEEE Trans. on Vehicular Technology,*, vol. 60, no. 5, pp. 2224–2238, 2011.

[87] Z. Li, J. Lin, M.-I. Akodjenou, G. Xie, M. A. Kaafar, Y. Jin, and G. Peng, "Watching videos from everywhere: A study of the pptv mobile vod system," in *Proc. ACM IMC*, 2012.

[88] J. Liu, X. Jiang, H. Nishiyama, and N. Kato, "Performance modeling for two-hop relay with node selfishness in delay tolerant networks," in *Proc. IEEE GreenCom*, 2011.

[89] R. Lu, X. Lin, T. H. Luan, X. Liang, X. Li, L. Chen, and X. Shen, "Prefilter: An efficient privacy-preserving relay filtering scheme for delay tolerant networks." in *Proc. IEEE INFOCOM*, 2012.

[90] A. McDiarmid, J. Irvine, S. Bell, and J. Banford, "CRAWDAD data set strath/nodobo (v. 2011-03-23)," Downloaded from http://crawdad.cs.dartmouth.edu/strath/nodobo, Mar. 2011.

[91] M. McNett and G. M. Voelker, "Access and mobility of wireless pda users," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 9, no. 2, pp. 40–55, Apr. 2005.

[92] A. Mei, G. Morabito, P. Santi, and J. Stefa, "Social-aware stateless forwarding in pocket switched networks," in *Proc. IEEE INFOCOM*, 2011.

[93] A. Mei and J. Stefa, "Swim: A simple model to generate small mobile worlds," in *INFOCOM 2009, IEEE*, April 2009.

[94] ——, "Give2get: Forwarding in social mobile wireless networks of selfish individuals," *IEEE Trans. on Dependable and Secure Computing,*, vol. 9, no. 4, pp. 569–582, july-aug. 2012.

[95] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Structures & Algorithms*, vol. 6, no. 2-3, pp. 161–180, 1995.

[96] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, "Epidemic outbreaks in complex heterogeneous networks," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 26, pp. 521–529, 2002.

[97] M. Musolesi and C. Mascolo, "Designing mobility models based on social network theory," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 11, no. 3, pp. 59–70, Jul. 2007.

[98] S. Nelson, M. Bakht, and R. Kravets, "Encounter-based routing in dtns," in *Proc. IEEE INFOCOM*, 2009.

[99] F. Neves dos Santos, B. Ertl, C. Barakat, T. Spyropoulos, and T. Turletti, "Cedo: Content-centric dissemination algorithm for delay-tolerant networks," in *Proc. ACM MSWiM*, 2013.

[100] M. E. J. Newman, "Spread of epidemic disease on networks," *Phys. Rev. E*, vol. 66, Jul 2002.

[101] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physical Review E*, vol. 64, no. 2, Aug. 2001.

[102] M. Newman, *Networks: An Introduction*. New York, NY, USA: Oxford University Press, Inc., 2010.

[103] G. W. Oehlert, "A note on the delta method," *The American Statistician*, vol. 46, no. 1, pp. 27–29, 1992.

[104] J. Ott, E. Hyytiä, P. Lassila, J. Kangasharju, and S. Santra, "Floating content for probabilistic information sharing," *Pervasive Mob. Comput.*, vol. 7, no. 6, pp. 671–689, Dec. 2011.

[105] J. Ott and J. Kangasharju, "Opportunistic content sharing applications," in *Proc. ACM NoM Workshop*, 2012.

[106] A. Panagakis, A. Vaios, and L. Stavrakakis, "On the effects of cooperation in DTNs," in *Proc. IEEE COMSWARE*. IEEE, 2007.

[107] A. Passarella and M. Conti, "Analysis of individual pair and aggregate intercontact times in heterogeneous opportunistic networks," *IEEE Trans. on Mobile Computing,*, vol. 12, no. 12, pp. 2483–2495, 2013.

[108] A. Passarella, M. Conti, C. Boldrini, and R. I. Dunbar, "Modelling inter-contact times in social pervasive networks," in *Proc. ACM MSWiM'11*.

[109] A. Passarella, R. I. Dunbar, M. Conti, and F. Pezzoni, "Ego network models for future internet social networking environments," *Computer Communications*, vol. 35, no. 18, pp. 2201–2217, 2012.

[110] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks," *Comm. Mag., IEEE*, vol. 44, no. 11, pp. 134–141, Nov. 2006.

[111] A. Pentland, R. Fletcher, and A. Hasson, "Daknet: rethinking connectivity in developing nations," *Computer*, vol. 37, no. 1, pp. 78–83, Jan 2004.

[112] A. Picu and T. Spyropoulos, "Dtn-meteo: Forecasting the performance of dtn protocols under heterogeneous mobility," *IEEE/ACM Trans. on Networking*, vol. PP, no. 99, 2014.

[113] A. Picu, T. Spyropoulos, and T. Hossmann, "An analysis of the information spreading delay in heterogeneous mobility dtns," in *Proc. IEEE WoWMoM*, 2012.

[114] A. Picu and T. Spyropoulos, "Forecasting dtn performance under heterogeneous mobility: The case of limited replication," in *Proc. IEEE SECON*, 2012.

[115] A.-K. Pietilainen and C. Diot, "CRAWDAD data set thlab/sigcomm2009 (v. 2012-07-15)," Downloaded from http://crawdad.cs.dartmouth.edu/thlab/sigcomm2009, Jul. 2012.

[116] A.-K. Pietiläinen, E. Oliver, J. LeBrun, G. Varghese, and C. Diot, "Mobiclique: middleware for mobile social networking," in *ACM Proc. WOSN*, 2009.

[117] A.-K. Pietilänen and C. Diot, "Dissemination in opportunistic social networks: the role of temporal communities," in *Proc. ACM MobiHoc*, 2012.

[118] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD data set epfl/mobility (v. 2009-02-24)," Downloaded from http://crawdad.cs.dartmouth.edu/epfl/mobility.

[119] M. Pitkänen, T. Kärkkäinen, and et al., "SCAMPI: service platform for social aware mobile and pervasive computing," *ACM Comput. Commun. Rev.*, vol. 42, no. 4, pp. 503–508, Sep. 2012.

[120] K. Poularakis, G. Iosifidis, A. Argyriou, and L. Tassiulas, "Video delivery over heterogeneous cellular networks: Optimizing cost and performance," in *Proc. IEEE INFOCOM*, 2014.

[121] S. M. Ross, *Introduction to Probability Models*, 9th ed. Academic Press, Elsevier, 2007.

[122] U. Sadiq, M. Kumar, A. Passarella, and M. Conti, "Service composition in opportunistic networks: A load and mobility aware solution," *IEEE Transactions on Computers*, vol. PP, no. 99, 2014.

[123] M. Satyanarayanan, "Mobile computing: The next decade," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 15, no. 2, pp. 2–10, 2011.

[124] V. Sciancalepore, D. Giustiniano, A. Banchs, and A. Picu, "Offloading cellular traffic through opportunistic communications: Analysis and optimization," *arXiv*, vol. 1405.3548, 2014.

[125] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, "CRAWDAD data set cambridge/haggle (v. 2009-05-29)," Downloaded from http://crawdad.cs.dartmouth.edu/cambridge/haggle, May 2009.

[126] K. Shanmugam, N. Golrezaei, A. Dimakis, A. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, Dec 2013.

[127] U. Shevade, H. H. Song, L. Qiu, and Y. Zhang, "Incentive-aware routing in dtns," in *Proc. IEEE ICNP*, 2008.

[128] T. Small and Z. J. Haas, "The shared wireless infostation model: a new ad hoc networking paradigm (or where there is a whale, there is a way)," in *Proc. of ACM MobiHoc '03*, 2003, pp. 233–244. [Online]. Available: http://doi.acm.org/10.1145/778415.778443

[129] T. Spyropoulos, K. Psounis, and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: the multiple-copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, 2008.

[130] ——, "Efficient routing in intermittently connected mobile networks: the single-copy case," *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 63–76, Feb. 2008.

[131] T. Spyropoulos, R. N. Rais, T. Turletti, K. Obraczka, and A. Vasilakos, "Routing for disruption tolerant networks: taxonomy and design," *Wirel. Netw.*, vol. 16, no. 8, pp. 2349–2370, Nov. 2010.

[132] T. Spyropoulos, T. Turletti, and K. Obraczka, "Routing in delay-tolerant networks comprising heterogeneous node populations," *IEEE Transactions on Mobile Computing*, vol. 8, pp. 1132–1147, 2009.

[133] P. Stuedi, I. Mohomed, and D. Terry, "Wherestore: Location-based data storage for mobile devices interacting with the cloud," in *Proc. ACM MCS Workshop*, 2010.

[134] K. Suh, C. Diot, J. Kurose, L. Massoulie, C. Neumann, D. Towsley, and M. Varvello, "Push-to-peer video-on-demand system: Design and evaluation," *Selected Areas in Communications, IEEE Journal on*, vol. 25, no. 9, pp. 1706–1716, December 2007.

[135] I. Trestian, S. Ranjan, A. Kuzmanovic, and A. Nucci, "Measuring serendipity: Connecting people, locations and interests in a mobile 3g network," in *Proc. ACM IMC*, 2009.

[136] uepaa, www.uepaa.ch.

[137] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Duke University, Tech. Rep. CS-200006, 2000.

[138] D. Wang, D. Pedreschi, C. Song, F. Giannotti, and A.-L. Barabasi, "Human mobility, social ties, and link prediction," in *Proc. ACM KDD*, 2011.

[139] X. Wang, M. Chen, Z. Han, T. Kwon, and Y. Choi, "Content dissemination by pushing and sharing in mobile cellular networks: An analytical study," in *Proc. IEEE MASS*, 2012.

[140] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic spreading in real networks: An eigenvalue viewpoint." in *SRDS*. IEEE Computer Society, 2003, pp. 25–34.

[141] J. Whitbeck, M. Amorim, Y. Lopez, J. Leguay, and V. Conan, "Relieving the wireless infrastructure: When opportunistic networks meet guaranteed delays," in *Proc. IEEE WoWMoM*, 2011.

[142] E. Yoneki, P. Hui, S. Chan, and J. Crowcroft, "A socio-aware overlay for publish/subscribe communication in delay tolerant networks," in *Proc. ACM MSWiM*, 2007.

[143] X. Zhang, G. Neglia, J. Kurose, and D. Towsley, "Performance modeling of epidemic routing," *Computer Networks*, vol. 51, no. 10, pp. 2867–2891, 2007.

[144] W. X. Zhou, D. Sornette, R. A. Hill, and R. I. M. Dunbar, "Discrete hierarchical organization of social group sizes," *Proc. Royal Society B: Biological Sciences*, vol. 272, no. 1561, pp. 439–444, Feb. 2005.

[145] H. Zhu, X. Lin, R. Lu, Y. Fan, and X. Shen, "Smart: A secure multilayer credit-based incentive scheme for delay-tolerant networks," *IEEE Trans. on Vehicular Technology*, vol. 58, no. 8, pp. 4628–4639, oct. 2009.

[146] Y. Zhu, B. Xu, X. Shi, and Y. Wang, "A survey of social-based routing in delay tolerant networks: Positive and negative social effects," *Communications Surveys Tutorials, IEEE*, vol. PP, no. 99, pp. 1–15, 2012.

[147] G. Zyba, G. Voelker, S. Ioannidis, and C. Diot, "Dissemination in opportunistic mobile ad-hoc networks: The power of the crowd," in *Proc. IEEE INFOCOM*, 2011.