FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

IMPROVEMENT OF THE ROUTING IN OPPORTUNISTIC NETWORKS BY THE APPLICATION OF UNSUPERVISED AND SUPERVISED MACHINE LEARNING TECHNIQUES

by

LADISLAVA SMÍTKOVÁ JANKŮ

A dissertation thesis submitted to
the Faculty of Information Technology, Czech Technical University in Prague,
in partial fulfilment of the requirements for the degree of Doctor.

Dissertation degree study programme: Informatics
Department of Digital Design

Prague, January 2019

Supervisor:
doc. Ing. Kateřina Hyniová, CSc.
Department of Digital Design
Faculty of Information Technology
Czech Technical University in Prague
Thákurova 9
160 00 Prague 6
Czech Republic

Abstract and contributions

The dissertation thesis deals with the issue of special routing algorithms designed for communication in opportunistic networks. The opportunistic networks (OPN) are networks disseminating messages with the store-carry-forward routing principle. The opportunistic communication networks are the ad-hoc networks where no assumption is made on the existence of a complete physical path between two nodes wishing to communicate. In opportunistic networks, the messages are transmitted when the node opportunistically meets another node; the characteristics of node movement can improve message transmission. The key function of OPN routing protocols is to make decisions on message forwarding. We proposed four routing schemes in this chapter: i) Hierarchical Routing with Clustering 1 (HRC1), ii) Hierarchical Routing with Clustering 2 (HRC2), iii) SVM-based routing, iv) Routing scheme combining GMRF (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm). The performance of the proposed method was tested on five simulation scenarios and compared to four well-known routing protocols as Epidemic routing with the limited message buffer, PRoPHET, First Contact and BUBBLE-Rap.

In particular, the main contributions of the dissertation thesis are as follows:

1. We proposed the routing protocol Hierarchical Routing with Clustering HRC1, which combines three strategies in order to improve routing in OPNs: i) the node affiliation with detected OPN geographic sector + use of *the sets of the detected geographic sectors*, ii) the node affiliation with the communication community constructed in spatio-temporal domain with time constraints, iii) epidemic routing.

2. We proposed the routing protocol Hierarchical Routing with Clustering HRC2 , which combines three strategies in order to improve routing in OPNs: i) the node affiliation with detected OPN geographic sector + use of the *the graph of the geographic sectors*, ii) the node affiliation with the communication community constructed in spatio-temporal domain with time constraints, iii) epidemic routing.

3. We proposed how use the Support Vector Machines to make decisions about routing in OPNs with regular node mobility patterns. We proposed SVM-based routing protocol, which uses an array of SVM classifiers to make decisions on routing.

4. We proposed Routing based on GMRF (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm).

Keywords:
Routing Protocol, Context-aware routing, Human mobility models, Clustering, Support Vector Machine, Opportunistic Network, Delay-Tolerant Network, Gaussian Markov Random Field.

Abstrakt

Disertační práce je věnována problematice směrování v oportunistických komunikačních sítích. Oportunistické komunikační sítě (OPN) jsou ad-hoc sítě, ve kterých se k šíření zpráv užívá schéma store-carry-forward. V těchto sítích se nepředpokládá, že by mezi dvěma uzly, kteří chtějí komunikovat, existovala v daném časovém okamžiku fyzická komunikační cesta. V oportunistických sítích jsou zprávy předávány tehdy, když se uzel dostane do komunikační vzdálenosti jiného uzlu. Klíčovou funkcí směrovacích protokolů OPN je tvorba rozhodnutí o předávání zpráv. V práci prezentuji čtyři směrovací algoritmy, které jsme navrhla:
i) Hierarchické směrování využívající shlukovou analýzu 1 (HRC1),
ii) Hierarchické směrování využívající shlukovou analýzu 2 (HRC2),
iii) Směrování s využití Support Vector Machine
iv) Směrování kombinující GMRF (Gaussian Random Fields) a ANMA (Active Node Movement Algorithm). Navržené metody byly testovány na pěti simulačních scénářích a porovnávány se čtyřmi dobře známými směrovacími protokoly jako Epidemické směrování s omezeným bufferem zpráv, PRoPHET, First Contact a BUBBLE-Rap a dosáhli v průměru výrazně lepších výsledků než stávající metody.

Hlavními přínosy disertační práce jsou zejména:

1. Navržený směrovací protokol Hierarchical Routing with Clustering HRC1, který kombinuje tři strategie pro zlepšení směrování v OPN.

2. Navržený směrovací protokol Hierarchical Routing with Clustering HRC2, který kombinuje tři strategie za účelem zlepšení směrování v OPN.

3. Navržený směrovací protokol SVM-based Routing který používá Support Vector Machines k rozhodování o směrování.

4. Navržený směrovací protokol GMR-ANMA Routing (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm).

Acknowledgements

I would like to express my gratitude to my dissertation thesis supervisor, doc. Ing. Kateřina Hyniová, CSc. She has been a constant source of encouragement and insight during my research and helped me with numerous problems and professional advancements.

Special thanks go to the staff of the Department of Digital Design, who maintained a pleasant and flexible environment for my research. I would like to express special thanks to the department management for providing most of the funding for my research.

Finally, my greatest thanks go to my family members, for their infinite patience and care.

Contents

List of Figures

List of Tables

Abbreviations

OPN	Opportunistic Network
DTN	Delay-Tolerant Network
MSG	Message
GMRF	Gaussian Markov Random Field
MANET	Mobile Ad-Hoc Network
VANET	Vehicular Ad-Hoc Network
OPPNET	Opportunistic Network
AODV	Ad-Hoc On Demand Distance Vector Routing
PROPHET	Probabilistic Routing Protocol using History of Encounters and Transitivity
PDSDV	Destination-Sequenced Distance Vector Routing
WAHN	Wireless Ad-hoc Networks
LS, LSE	Least Squares Estimate
MLE	Maximum Likelihood Estimate
AR Model	Autoregressive Model
SVM	Support Vector Machine

CHAPTER 1

Introduction

1.1 Motivation

The opportunistic communication networks are the ad-hoc networks where no assumption is made on the existence of a complete physical path between two nodes wishing to communicate [152]; the source and destination nodes needn't be connected to the same network at the same time. This assumption makes the routing in these networks difficult.

In contrast to the common ad-hoc networks with connected topology, in opportunistic networks, the source and destination nodes needn't be connected to the same network at the same time, but they are allowed to exchange messages between them due to techniques of the opportunistic networking. These techniques allow nodes to connect and disconnect anytime. In opportunistic networking, no limitations are also set on the nodes to keep their positions; the nodes can move. This opportunity networking paradigm opens a space for a number of novel application scenarios.

The general concept of opportunistic ad-hoc network does not require at one time there was a complete communication path between a source and a destination node. It is not required a graph describing mathematically general topology of the opportunistic ad-hoc network to be a connected graph.

In opportunistic networks, the messages are transmitted when the node opportunistically meets the another node; the characteristics of node movement can improve message transmission. In practice, the network nodes can be mobile robots, wireless equipment carried by people, vehicles, wild animals, unmanned aerial vehicles.

Opportunistic networks (OPN) are networks disseminating messages with the "store-carry-forward" routing principle. The key function of OPN routing protocols is to make decisions on message forwarding. The routing metrics are designed in order to select the most optimal nodes which have the highest probability to be a part of the paths of successfully delivered messages with respect to maximization of message delivery ratio and the minimization of overhead cost and message delivery delay.

1.2 Problem Statement

The motivation for this research is the fact that despite gains in the area of opportunistic networking and appearance of many sophisticated solutions for routing in OPNs, the general perception of the problem is that the application of deep learning methods is still very abstract and there is still space for research. As the machine-learning based data driven individualised models of human mobility appear in traffic analysis and planning, the routing in OPNs is possible to further improve. With this goal, this work seeks solutions for applications of machine learning in opportunistic networking.

There are three explicit objectives of our research:

1. Problem: to propose an enhancement of routing algorithms using unsupervised learning and examine the impact of this enhancement on communication in opportunistic networks. We want to verify or rebut the hypothesis that it is possible to extract the knowledge from these data using unsupervised learning and use this knowledge to improve a routing algorithm in such a way the delivery time of messages are shortened and an overall communication in the opportunistic network is improved in comparison to standard routing methods. Experimental validation of the method will be performed on the simulated data in the opportunistic network simulator.

2. Problem: to propose an enhancement of routing algorithms using supervised learning and examine the impact of this enhancement on communication in opportunistic networks.

If there are observable motion patterns in motion of nodes which repeat during the time, it is likely that nodes will meet some nodes more often than the others. It can be assumed that some places on the routes are more useful for passing the message than others from the viewpoint on message delivery time. So we can collect data about the time and place of the passing messages, the nodes that receive message and the time when the messages were delivered to the destination node.

We want to verify or rebut the hypothesis that it is possible to extract the knowledge from these data using supervised machine learning method and use this knowledge to improve a routing algorithm in such a way the delivery time of messages are shortened and an overall communication in the opportunistic network is improved in comparison to standard routing methods.

3. Problem: to propose an enhancement of routing algorithms using unsupervised learning (statistical node mobility models) and the active node behavior and examine the impact of this approach on communication in opportunistic networks. The active node behavior means that the node itself actively changes its route in order to get to the location more suitable for forwarding messages.

We want to verify or rebut the hypothesis that active node behavior and active decision about a place of message passing can influence the delivery time of this To study the implementation of statistical node mobility models and the active node behavior in routing algorithms for opportunistic networks and to confirm the hypothesis that the active node behavior in opportunistic network supported by the appropriate statistical node mobility model can improve routing and communication in the opportunistic network.

1.3 Related Work/Previous Results

The success of the routing in opportunistic networks depends on the knowledge of the network topology deployment in a near future. The routing effectiveness increases, if the routing process is capable to predict the changes in the network topology [82]. In recent years, the research in OPN routing has attracted much attention because of the wide range of mobile applications of OPN routing. Besides the reactive routing methods, in which the nodes compute forwarding strategies through the contact history, without a global or predetermined knowledge, also proactive routing becomes popular. The examples of proactive routing protocols are knowledge-based routing schemes [82], RAPID [6], Routing in cyclic mobile space [115], Capacity-aware routing using throw-boxes [63], and Mobyspace [103] or ML-SOL [186]. The examples of reactive routing protocols are First Contact, Epidemic[206], PROPHET[114], Spray and wait [191], Seek and focus[193], Spray and focus [192], Bubble Rap [76], Social network-based multi-casting [58], or Island Hopping[173]. Also social-aware routing received particular attention and a lot of community based OPN routing protocols which uses knowledge obtained from community detection and formation in order to improve routing performance have been proposed. Examples of community-based routing protocols include BUBBLE RAP [76], LocalCom, Gently and Diverse Routing, ML-SOR. The main idea of community-based routing is that relationship of the users is reasonable information for predicting future contact opportunities. The community has a strong impact on human mobility pattern. The community-based routing schemes consists of two phases. In the recent years, the particular attention to the analysis of communities within networks was given in various disciplines, particularly but not only in mathematics, physics and biology. Scientist have become interested in the study of networks describing topologies of wide variety real systems [47]. Biochemical networks, social networks, communication networks, transportation networks, text databases networks, world wide we and much more. Multi-layer approaches to routing in OPN model the OPN as a structure of mutually connected layers. In addition to contact graphs describing physical encounters of nodes, they use social layers reflecting the real world contacts. The presented multi-layer social graph is based on the social graphs extracted from different sources such as Facebook, Twitter or e-mail communication. The examples are ML-SOL, Social Role Routing or MobiClique. However, to the author best knowledge, there are only few publications available in the literature that discuss application of machine learning based models of human mobility in the area of opportunistic networking, although in the field of human mobility modeling using data mining a lot of scientific papers have been published. Bazzani et al. [10] presented statistical analysis of a mobility dataset obtained in the Florence urban area. They tested by the probability distribution and the moving object activity of robust statistical laws. Due to the rapid development of wireless technologies and high-tech applications, many publications have appeared in recent years, which are related to analysis of large datasets. Liu et al. [116] analyzed data consisting of 85 million GPS points of taxicabs collected in Wuhan, China. They proposed mobility model based on spatio-temporal paths of moving nodes and spatio-temporal clustering algorithm, which uses spatial clustering of node positions at different times and a method of complex hull to merge these clusters into spatio-temporal ones. Hoque et al.[71] analyzed GPS data of taxicabs obtained in the San Francisco area by application of clustering and statistical methods. Cheng et Anbaroglu et al. [40] proposed a spatio-temporal clustering algorithm for complex temporal networks analysis in spatial, temporal and thematic domains and tested it on data obtained from a part of London's traffic network. Schneider et al.[175] proposed the application of network motifs in human mobility analysis. They constructed daily human mobility networks from CDR data for Paris over a period of 6 months and from travel survey data for Paris for one day. They reported, that they identified 17 unique motifs. Jiang et al. [84] have applied a similar approach to extract human daily motifs. They have constructed daily human mobility networks from triangulated mobile phone CDR data for one million users in Boston. They have reported similar findings. Furthermore, they have proposed a probabilistic inference method to use motifs, time of day, activity sequence, and land use related information to further infer activity types and traffic patterns Widhalm et al. [214] proposed methods for inferring human activity types from data extracted mobile phone data and land use data for the cities of Boston and Vienna.

1.4 Goals of the Dissertation Thesis

1. Propose the routing metric combining utilization of geographical data and unsupervised machine learning (cluster analysis).
2. Propose the routing metric which uses supervised machine learning technique as a part of decision making mechanism.
3. Propose the routing method, which continuously evaluates the network state and can enhance the routing process by active changes in node behaviour.
4. Select the appropriate performance metrics and prepare simulation scenarios for evaluation of the proposed methods.
5. Experimental evaluation of all proposed routing methods in simulation environment.

1.5 Structure of the Dissertation Thesis

The dissertation thesis is organized into five chapters as follows:

1. *Chapter 1 "Introduction"* consists of four parts: motivation, problem definition, a short overview of the previous work and setting the main objectives. We described

the motivation behind our efforts, determined three mutually independent problems to be solved and set the main objectives of the of the dissertation thesis.

2. *Chapter 2 “State-of-the-Art”* introduces the state of the art of the opportunistic networking and surveys.

3. *Chapter 3 “Overview of Our Approach”* deals with the proposed methods to solve the problems.

4. *Chapter 4 “Main Results”* describes the results of experimental verification of the proposed methods in the simulation environment.

5. *Chapter 5 “Conclusions”* summarizes the results of our research, provides the directions for further research.

Chapter **2**

Background and State-of-the-Art

2.1 Theoretical Background: Introduction to Ad-hoc Networks

We start with the concept of the ad-hoc network. The phrase ad-hoc comes from Latin. The Webster's dictionary [124] or, as defined in [216], the meaning of this word phrase ad hoc follows: "...only in this case, the purpose of this rarity." The equivalent to the phrase ad-hoc network could thus be a phrase *a network built according to the actual needs*.

The ad-hoc network is defined as a decentralised network, which satisfies the following conditions: i) the network does not rely on a pre-existing infrastructure, rather it is formed *on demand* ii) peer-to-peer communication among the nodes.

Each node participates in routing by forwarding packets addressed to other nodes. The communication path is created dynamically on the basis of network connectivity and with the respect to implemented routing algorithm. the definition of ad-hoc network does not include any constraints set to communication medium. In recent years, the wireless ad-hoc networks (WAHN or recently proposed abbreviation WANET) have attracted a lot of attention.

Another important feature of the common ad-hoc network is the lack of requirement of an existence of a direct communication path between any pair of the nodes, see Fig. 2.1. If the node A communicates with the node B, it can make via a communication path including nodes C and D. In computer literature, it is sometimes incorrectly reported that in ad-hoc networks consisting of computers must be nodes (individual computers) in range of each other so that everyone can communicate with everyone. If this constraint appear it is related to the technology in use. It is not a general property of ad-hoc networks. This technology limitation is often incorrectly generalized on all ad-hoc networks.

The general concept of ad-hoc network does not require the establishment of complete communication path between a source and a destination node using a communication media. Instead of that, the messages can be disseminated with the "store-carry-forward" routing principle. A static graph describing mathematically an immediate topology of

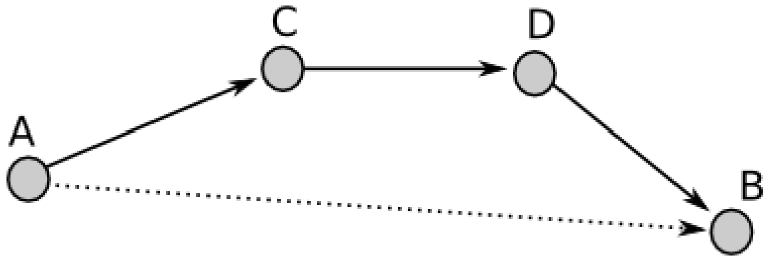

Figure 2.1: Example of communication in the ad-hoc network: the node A communicates with node B, it can do so in the example via a communication route comprising nodes C and D.

the ad-hoc network is not required to be a connected graph. An example of this kind of network is shown on Fig. 2.2. In this network, the source node A sends a message to the destination node F. The message is captured by the node C, which forwards it to the node D. The node D moves and once it gets into the communication range of the node E, it forwards the message to the node E. The node E forwards the message to the destination node F.

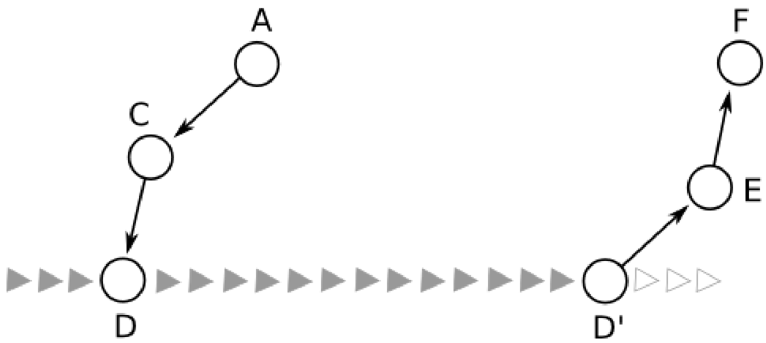

Figure 2.2: Example of opportunistic communication in an opportunistic ad-hoc network.

A network that supports this method of message delivery is called an **opportunistic ad-hoc network (opportunistic network)**. Due to the wide application possibilities of opportunistic networks, communication protocols and behavior of these networks are the subject of intense research since the the concept of opportunistic ad-hoc networking was introduced in 2006 [152].

Despite the interesting properties of the general ad-hoc networks, the research has been limited for many years on the special cases of ad-hoc networks. The ad-hoc networks, which had the connected topology (connected graph) have been studied. Depending on whether the system allowing unidirectional or bidirectional communication, the network topology was described by strongly or weakly connected graph. In ad-hoc networks of this type, where the ad-hoc network is formed by connecting the communicating nodes in the network at first, and then started the communication between nodes, which have been already connected, the research was focused on the communication protocols, routing algorithms, algorithms that prevent network congestion and analysis of the properties these networks. Although ad-hoc networks of this type represent only a subset of ad-hoc networks, most publications often use the term ad-hoc networks to describe these types of networks due to lack of more precise terminology. To distinguish this type of networks from the general ad-hoc networks, we introduce a more precise term. We call them the **ad-hoc networks with connected topology**.

2.2 Wireless ad-hoc networks with connected topology

In wireless ad-hoc networks with connected topology, it is supposed i) to create a communication path between the source and the destination node, ii) ensure the existence of this route for some time. Communication in these networks consists of two steps:

(a) a search for an existing route between the source node and the destination one; and

(b) sending messages through this route.

The existence of the route is monitored in some communication protocols. If the route is destroyed or the transmission characteristics of the route changes significantly, new communication path connecting the source and the destination node must be established. Routing algorithms for these types of networks are usually designed to minimize or maximize some predetermined criterion, such as minimizing the number of sent messages, minimizing the total energy consumption or maximizing the lifetime of the network.

2.2.1 Routing in ad-hoc networks with continuous topology

2.2.1.1 Reactive Routing Protocols

Nodes in ad-hoc networks with reactive communication protocols utilize so-called *source routing*. The routing initialization process is then usually carried by flood plain algorithm. During initialization, the transmission path connection source and destination nodes is found. The messages are sent through the found transmission path. Routing algorithms that belong to this group differ from each other mainly by optimizing the initialization phase.

The routing protocol *Ad-Hoc On Demand Distance Vector Routing (AODV)* [153] is a reactive modification of a routing protocol DSDV. Routing information are not created by default on all nodes of ad-hoc network, but only at the request of one of them. If the source node wants to establish a communication connection to the node, which it has no information on except its address, or, if the transmission path to the destination node has already expired, the source node sends a route request message. This report has a unique identifier. This message is sent as a broadcast. Any node, which receives a route request message and which is not the destination node of this message at the same time, analyses its routing table stored locally in the node. If the node does not know the path to destination node, it adjusted the message and sends it as a broadcast. At the same time, the node creates a reverse entry in the routing table. If the node knows the path to the destination node, it compares the sequence number of the target contained in the report with a stored number. If the message number is higher or equal than the stored one, and simultaneously has a better metric, the node as it did not know the way, if not directly respond route reply message to the sender (after updating metric in the routing table based on a comparison with incoming metric). If it receives a route request message, logs the direction from which the request came with the smallest metric, and sends a route reply. It has already spread engulfing.

To detect a transmission path interruption, a method based on sending so called “hello” messages is used. All nodes transmit at a certain time interval a “hello” message to the neighboring nodes. Each node maintains a list of neighboring nodes. If any of the neighboring node does not send a response, the node, which sent the message “hello”, marks this non-responding node as an unplugged one. If this non-responding node is a part of a transmission path, the node, which identified the transmission path breakup, sends a special route reply message to the other nodes on the path. The report is gradually transferred to the destination node; references in the tables are stored in the nodes forming the path are adequately updated and finally, the transmission path is complete again. In comparison to the routing protocol DSDV, the AODV routing protocol seems to be very effective. Each node maintains only the information, which it needs to operate as a source node, a destination one or a node on-the-path. The path between the source and the destination node is found very quickly, mainly due to the dissemination route request and route reply messages. In comparison to the routing protocol DSDV, any implementation of the routing protocol AODV is difficult and challenging task.

The routing protocol *Simplified Ad-Hoc On Demand Distance Vector Routing (AODVjr)* [28] is a simplified version of the AODV routing protocol. A route request message has a simpler structure and the source node disseminates this message using a flood algorithm. The node on-the-path responds only to the first route request message with the specified identification number, other reports are thrown away. It is assumed, the node from which the route request message arrives first is also the most convenient node for sending the response, respectively. create a transmission path. The destination node corresponds to the route reply message. In order to detect a path interruption, the maintenance of an active bidirectional communication connection between the source and the destination nodes is necessary. If the connection is unilateral, the destination node performs the path validation using “connect” messages. AODVjr is a quite elegant method. It preserves all the advantages of AODV, but it is less difficult to implement. The disadvantage of both algorithms is the inability to respond to the improvement of the conditions for the transfer. If there is functional transmission path between source and destination nodes, both the algorithms do not respond to the emergence of new routes between source and destination nodes, even if they have better performance. Data are forwarded along the found transmission path as long as this path exists.

The routing protocol *Dynamic Source Routing in Ad Hoc Wireless Networks (DSR)* [85] uses so-called source routing. Source routing is initiated by a node that wishes to establish a connection. This node sends the route request packets that are identified by a unique identifier and are disseminated using flood plain method. Each network node through which the route request packet is transmitted, placed its record the route request packet. The destination node responds to the route request packet with route reply. This message is usually connected to the next route request requirement to create return path. If the destination node becomes unavailable, the node using the routing protocol cancels the corresponding entry in the routing table. The entry connected to the path in which the node is unavailable is canceled. Then the node sends the route error message that informs the other nodes in the transmission path. The source node must then ask for a transmission path initialization again.

The routing protocol *Temporary-Ordered Routing Algorithm (TORA)* proposed in [148] is based on the ideal of maintaining an acyclic graph of all possible transmission paths in each network node. Developing and maintaining these graphs are not cost-effective. This routing protocol puts emphasis on minimizing the number of transmitted messages through the network. The edge of the graph is therefore initialized only then, if it is not exist in a model. The node itself responds to the edge interruption only and only if the interrupted edge is irreplaceable. It is not necessary for each problem to perform the complete initialization of the graph representing the network topology. The response to topology change is very rapid. In networks using this routing protocol, the communication flow between source and destination nodes can be split to the messages transmitted via different routes.

2.2.1.2 Proactive routing protocols

Proactive routing protocols aggressively spread in the network specific predefined information. The network nodes create routing rules on the basis of this information. This group of routing protocols does not include only the routing protocols for ad-hoc networks, but also the protocols such as RIP or OSPF used on the Internet; these protocols are not suitable for the communication in ad-hoc networks, because they require the default address aggregation. If we use the routing protocol RIP or the routing protocol OSPF in ad-hoc network, nodes would need to exchange information about all other nodes. It would probably lead to a significant increase in overhead. Proactive protocols designed for ad-hoc networks are therefore often designed to realize an exchange of information needed to implement routing only among selected nodes.

An example of a proactive routing protocol might be a routing protocol *Highly Dynamic Destination-sequenced Distance Vector Routing (DSDV)* [154], which uses a distributed Bellman-Ford algorithm. It is based on transmission of a distance vector among the nodes in a regular time intervals. This distance vector contains a price to be payed to jump to the next node. The DSDV routing protocol is designed to prevent forming loops and to recognize changes in the network topology. Its advantages are simplicity, low overhead, low computational and memory requirements. The DSDV routing protocol can use different metrics for transmission path evaluation. Each node has stored a distance vector in its local memory (so called distance vector storage, DVS). At regular intervals, the node transmits to the network a part of the distance vector, known as DVB. At the same time, the node receives DVB from other nodes and processes. Lack of the DSDV routing protocol lies in the fact that this protocol does not solve the problem of aggregation; the nodes must remember the DVS data on all other nodes. Another weakness is the inability of the DSDV routing protocol to respond flexible enough to the dynamic changes in the network. Any change of a metrics needs a long time to be distributed across the network.

Another proactive routing protocol developed for ad-hoc networking is the routing protocol *Maximum Residual Packet Capacity (MPRC)*. MPRC routing protocol [7] is based on a strategy to maximize the time remaining until the collapse of the network. The cause of collapse of the network is disconnecting the nodes from this network due to exhaustion of energy resources. It is therefore a protocol for ad-hoc networks in which it is difficult to supplement energy supplies (for example a network for wild birds monitoring, a special network of sensors, networks, solar power, etc.). The information on node energy reserve is used in routing. Each node calculates a metric indicating the maximum number of bits that it is still possible to pass through this node before the energy reserve is exhausted. Then, the transmission path evaluation is calculated for each transmission path. The MPRC routing protocol selects the path that contains the strongest of the weakest nodes of all possible transmission paths. The MPRC routing protocol is able to select paths that do not include nodes with the smallest energy resources. The weaknesses of this protocol include long initialization time and a slow response to changes in the network.

2.2.1.3 Routing protocols based on a backbone topology

This group of routing protocols have been proposed for WAHN networks. Based on the observation, that some of the technological implementations of nodes in the WAHN networks have higher energy consumption when they are switched to the reception mode than when they are found in an idle mode. From the viewpoint of energy resources preserving, it is preferred to maintain the nodes of the network in idle mode as long as possible. This is achieved by introducing a set of selected nodes, called coordination nodes, which remain in a reception mode, while the other ones are switched to the idle mode. The node in the idle node is not able to receive messages. The set of the coordination nodes must be chosen so as to cover all nodes of the network. The function of a "coordination node" rotates through all the nodes of the network. An example might be the routing protocol *Span* [36], which provides

(a) a network coverage by coordination nodes,

(b) the rotation function of the coordinating nodes and

(c) minimizing the number of nodes selected to be coordinating nodes without impacting on network capacity or transmission delay.

The computation of coordination nodes is implemented locally (without central control). The computation of coordination nodes is repeated at regular time intervals. Node, which wishes to become a coordination node, reports that during the election period to other nodes with a certain delay from the beginning of the electoral period. The delay is calculated on the basis of selected parameters, e.g., depends mainly on the number of neighboring stations and energy reserves of the node. It also has a random component. The delay affects the probability the node becomes a coordination node.

2.3 State-of-the-Art: MANET, VANET, Delay-tolerant Networks and Opportunistic Networks

Although the property of the node mobility is implicitly included in the properties of ad-hoc networks, the name of Mobile Ad-hoc Network (MANET) was introduced to emphasize the property of a node mobility and the node freedom in connecting to the network and disconnecting from the network during the communication.

The Vehicular Ad-hoc Networks (VANET) are special case of self-organizing MANET networks, which nodes are formed by moving vehicles. Their basic characteristics include: i) a movement of network nodes, ii) limited degree of freedom in the movement patterns of network nodes [67]. VANET networks represent a large and growing class of MANET networks.

The most important research DTN networks is centered around projects funded by DARPA: Disruption Tolerant Networking [137]. The Control-Based Mobile Ad-Hoc Networking (CBMANET) [121], Military Networking Project [227] Connection-less Networks [90] and around open project Delay Tolerant Networking Research Group [33], which has been designed as part of the Internet Research Task Force. The terms "Delay Tolerant Network" or "Delay and Disruption Tolerant Network" and "Opportunistic Network"are used interchangeably [29]; therefore these terms often used as equivalent terms. On the other hand, there is a group of authors [152], which highlights the difference between DTN networks and opportunistic networks, primarily the fact, that the concept of DTN networks, as introduced in [33], and communication protocols for these networks include also networks based on the same concept as Internet, while opportunistic networks are networks of mobile nodes disseminating messages with the "store-carry-forward" routing principle.

2.4 Routing in Opportunistic Networks

This section deals with routing in OPN/DTN networks disseminating messages with the "store-carry-forward" routing principle. Opportunistic routing on networks, where nodes do not fulfill this routing principle. In the following, we will use the terms routing and forwarding interchangeably.

2.4.1 OPN Routing Protocols Taxonomy

In recent years, many different algorithms for routing in OPN/DTN have been proposed. Unfortunately, there is no unique taxonomy of OPN routing protocols, instead of that, several taxonomies have been proposed. The taxonomies differ in accordance to which aspect of OPN routing their authors have preferred.

Jain et al. [82] have proposed the first taxonomy for routing in OPNs (called DTNs in Jain's paper). In order to set up the taxonomy, they have proposed so called *oracles*, which maintain the knowledge about the network. This model contains four different oracles: i) Contact Summary Oracle, ii) Contacts Oracle iii) Queuing Oracle and iv) Traffic Demand Oracle. In the point of view of Jain et al. [81], the key objective of research on routing-in-OPNs algorithms is understanding the relationship between algorithm performance and the use of these oracles. They have defined these oracles as follows: "Contacts Summary Oracle This oracle can answer questions about aggregate statistics of the contacts. In particular, the contacts summary oracle provides the average waiting time until the next contact for an edge. Thus, the contacts summary oracle can only respond with time-invariant or summary characteristics about contacts. Contacts Oracle This oracle can answer any question regarding contacts between two nodes at any point in time. This is the equivalent to knowing the time-varying DTN multi-graph. The contacts summary oracle can be constructed using the contacts oracle, but not vice versa. Queuing Oracle This oracle gives information about instantaneous buffer occupancies (queuing) at any node at any time, and can be used to route around congested nodes. Unlike the equivalent to knowing the time-varying DTN multi-graph. The contacts summary oracle can be constructed using the contacts oracle, but not vice versa. Queuing Oracle This oracle gives information about instantaneous buffer occupancies (queuing) at any node at any time, and can be used to route around congested nodes. Unlike the other oracles, the queuing oracle is affected by both new messages arriving in the system and the choices made by the routing algorithm itself. We expect it to be the most difficult oracle to realize in a distributed system. Traffic Demand Oracle. This oracle can answer any question regarding the present or future traffic demand. It is able to provide the set of messages injected into the system at any time. Unlike other oracles, the queuing oracle is affected by both new messages arriving in the system and the choices made by the routing algorithm itself. We expect it to be the most difficult oracle to realize in a distributed system. Traffic Demand Oracle This oracle can answer any question regarding the present or future traffic demand. It is able to provide the set of messages injected into the system at any time." (Jain, Fall & Patra, 2004) These oracle definitions are important, because they define outlines of the problem which each designer of the OPNs routing protocols meets. The proposed taxonomy classify routing protocols in the context of these oracles.

Pelusi et al. [152] have adopted a hierarchical taxonomy of OPN routing protocols proposed by Zhang [228]. At the highest level of this taxonomy, the OPN routing algorithms are classified into two classes: i) *routing without infrastructure* and ii) *routing with infrastructure*. The class first class contains methods designed for completely flat ad hoc networks, while the second class contains algorithms in which the some form of infrastructure is used in order to opportunistically forward messages. Research presented in this thesis is considered to be applicable only on OPNs which support the routing without infrastructure. The class of *routing without infrastructure* methods is divided to i) *dissemination-based routing*, which are some versions of controlled flooding, and ii) *context-based routing*. Context-based routing protocols usually do not adopt flooding schemes, but use knowledge of the context in order to identify the best node for message forwarding. In comparison to dissemination-based routing protocols, context-based routing protocols significantly reduce the number of duplicate messages in OPN. The disadvantage of context-based routing are higher values of message delivery delay.

Flood routing algorithms in OPNs work as follows: the source node sends a message, which is addressed to the destination node. The message is disseminated across the network using"flood" strategy. These algorithms are based on the idea that even if there is no knowledge about the location of the destination node or which of the available nodes should be the next step for sending the message, the message can be disseminated in the network and deliver to all nodes in the network. These algorithms are successful in opportunistic networks with high mobility of the nodes. Disadvantages are the high demands on resources and the high risk of a network congestion [151]. The *routing with infrastructure* algorithms are classified into two subgroups: i) ii) *Mobile Infrastructure Routing* or

Carrier Based Routing. In both cases the infrastructure is composed by special nodes that are more powerful with respect to the other nodes commonly present in the ad hoc network. Routing Protocols based on Fixed Infrastructure An example of a routing protocol with a fixed infrastructure is Infostation model [60]. This type of protocol is introduced for the networks, which contain special fixed nodes, which can be seen as a special form of base stations (called infostations), through which it is possible to send messages. This type of a network no longer represents pure ad-hoc network in accordance to the definition adopted in this work. If a node operating in a network communication protocol Infostation wishes to communicate with another node, it sends the message directly to the base station, if the base station is in its communication radius. Otherwise, it tries to send a message opportunistically through another node. In this concept, the base stations do not use the possibility of opportunistic routing, so the network uses a fixed routing infrastructure essentially. It is better to speak about the "hybrid network communication protocols" than the opportunistic networking. This is different for routing protocols in mobile infrastructure.

Mobile Infrastructure Routing or *Carrier Based Routing* use special nodes called mobile data collectors. These special nodes move in the network, and their directions may be random or predetermined. The mobile data collectors may be the only entities responsible for transmission of messages, if enabled message transfer only between mobile data collectors and ordinary node network, or they can strengthen connectivity in sparse networks and to guarantee that even isolated nodes will affect the communication process. An example of a communication system using this architecture and routing protocol is a MULE system described in [81], [229].

Hong et al. [219] have proposed a hierarchical taxonomy of OPN routing protocols based on routing protocol reactivity or pro-activity. At the highest level, the OPN routing protocols are classified into two categories: proactive routing and reactive routing protocols. The proactive routing class contains methods which use the centralized or offline knowledge about the mobile network to make the routing decision. The reactive routing class contains methods, in which the nodes compute forwarding strategies through the contact history, without a global or predetermined knowledge. The examples of proactive routing protocols are knowledge-based routing schemes [82], RAPID [6], Routing in cyclic mobile space [115], Capacity-aware routing using throw-boxes [63], and Mobyspace [103] or ML-SOL [186].The examples of reactive routing protocols are First Contact, Epidemic[206], PROPHET[114], Spray and wait [191], Seek and focus[193], Spray and focus [192], Bubble Rap [76], Social network-based multi-casting [58], or Island Hopping[173]. Context-based algorithms do not use flood techniques but they try to select the nodes to which the message should be forwarded.

Another approach to OPN routing protocols classification has been adopted by Moreira et al. [129], [130], [131], who have proposed a hierarchical taxonomy which is taking into account both way of message transmission and OPN social and topological features, such as contact frequency and age, resource utilization, community formation, common interests or node popularity. At the highest level, the OPN routing protocols are classified into three categories: i) forwarding-based routing protocols, ii) flooding-based routing protocols, and iii) replication-based routing protocols.

Xia et al. [221] proposed a hierarchical taxonomy of OPN routing protocols primary in the context of social aware routing. At the highest level, the OPN routing protocols are classified into two categories: i) unicast routing and ii) multicast routing. The unicast routing protocols are further divided into two groups: i) community-based routing and ii) community-independent routing. *Community-based routing* class of OPN routing protocols contains OPN routing protocols, which uses knowledge obtained from community detection and formation in order to improve routing performance. As examples of community-based routing protocols, Xia et al. [221] have discussed BUBBLE RAP [76], LocalCom, Gently and Diverse Routing. The community-independent routing protocols do not detect communities. They use the context information and social network metrics such as centrality or similarity.

Ahmad et al. [2] have proposed a taxonomy, which divides OPN routing protocols into six main classes: Geographic, Link State-aware, Context-aware, Probabilistic, Optimization Based and Cross Layer routing protocols. Geographic routing protocols are routing schemes, which use location data as context to make routing decisions. Although these methods have a context-aware character (location data = context), the scientific papers address the geographic routing protocols usually as a homogeneous independent group of routing protocols. As the context-aware routing, the routing schemes taking into account node mobility patterns, social community affiliation or node importance in network are understood. Cadger et al. [27] classified geographical routing schemes into four groups: i) greedy forwarding ii) face routing, iii) hybrid greedy face routing, iv) routing in 3D space. Greedy forwarding schema use a very simple routing metrics: the messages are forwarded to the neighbour node, which is located closest to the destination. The drawback of greedy forwarding consists in the risk of the genesis of loops where the messages are forwarded backwards. In order to avoid the genesis of loop, the technique of message drooping by nodes is used.

Routing protocols based on coding [215], [210], [38], [112], [167], [50] use merging multiple reports into one report using coding. One message is sent instead of a group of messages. The transmission channel then forms MIMO (multi-input multi-output) system. The reconstruction of coded messages by the node is performed using iterative algorithms that maximize the mutual independence between the outputs of the transmission channel. Most of these methods is based on the Principal Value Decomposition (PCA or Karhuen-Loeve Transform), and in particular on Independent Component Analysis (ICA). These methods are similar to those ones used in MIMO system in automatic control to find mutually independent variables in measured output variables. An excellent summary of the theoretical problems of methods derived from PCA and ICA and its application in MIMO systems is presented in [42]. Opportunistic coding methods were originally developed for the transmission of multimedia content. The details can be found in [181], [180].

2.4.2 Selected Routing Algorithms

This section describes several routing schemes in detail.

2.4.2.1 First Contact

First Contact routing protocol is a forward-based routing strategy; only one copy of a message exists at a certain point in time in a system. The node that carries a message, forwards this message to the first node, which it encounters, unless the message is transmitted to the destination node.

2.4.2.2 Epidemic Routing

Vahdat et al. [206] proposed Epidemic routing protocol for message transfer in OPNs. Epidemic routing is based on concept of complete flooding. Each node maintains two buffers in which first buffer is used for storing the messages generated by the nodes encounter, they compare the contents of their messages buffers. Each of them accepts the copies of messages, which are not contained is its buffer. After the node forwards a message, the copy of this message still remains in its buffer. If sufficient resources are available (message buffers and bandwidth capacity), Epidemic Routing has high message delivery ratio and low message delivery delay. This protocol has high demands on node message buffer capacity and bandwidth capacity.

2.4.2.3 Spray and Wait

The Spray and Wait protocol [191] provides an improvement over the Epidemic routing protocol by controlling the level of flooding. In this protocol there are two phases: the Spray phase and the Wait phase. As in Spray phase, every message originated at the source node is passed to L distinct relays in the network i.e. L copies of the message are spread over the network by the source code. If the destination was not found in the spray phase, then in Wait phase each relay node having a copy of message performs the direct transmission of the message to the destination itself. The performance of this protocol depends on the value of L, smaller the value of L makes it similar to Direct delivery protocol and larger the value of L makes it similar to the Epidemic protocol. This protocol has less number of transmissions and delay as compared to Epidemic Routing.

2.4.2.4 MV: Meeting and Visits Routing Protocol

The routing protocol *Routing Meeting and Visits* (MV Routing) published in [26] is a probabilistic routing protocol. The algorithm MV routing use the probabilistic evaluation of nodes. These probabilities are obtained experimentally and they reflect the history of node encounters and node location history.

2.4.2.5 PROPHET

The another variant of the epidemic routing protocol is the routing protocol called *Prophet (Probabilistic Routing Protocol using History of Encounters and Transitivity). Prophet* is based on the same principle as the MV Routing or the expansion of the epidemic routing protocol presented in [202]. *Prophet* uses an adaptive algorithm that computes a set of probabilities for successful delivery to known destinations in the opportunistic network. These probabilities are called delivery predictabilities. Each node maintains a table for delivery predictabilities to other nodes. If two node meets, the tables of delivery predictabilities are updated. The messages during opportunistic meeting of two nodes are passed only if the node has no better possibility to pass to message in accordance to the table of delivery predictabilities. The current version of the PROPHET protocol is documented in RFC 6693 [78]. Stochastic analysis of this type of routing can then be found in [113], [114]. More variations of epidemic routing can be found in [165], [209].

2.4.2.6 LABEL

Hui et al. have proposed a routing scheme called LABEL; the proposed method is based on labeling nodes in accordance to their affiliation into groups and simple label-based message transmission rule. If two nodes encounter, the message is forwarded only if the destination node is from the same group as the encountered node. Hui et. al have tested this method on small real-world human mobility datasets. The method improves routing performance on small OPNs with well-connected nodes, but fails in more structured environments, where the destination nodes are socially far from the source node. However, this method is usually perceived as the first attempt to differentiate among nodes from the viewpoint of their social activity, and use this information in order to improve routing performance.

2.4.2.7 CAR: Context Aware Routing

Context-aware Routing (CAR) was introduced in [135], [133]. It is based on probabilistic evaluations of other nodes. Each node maintains a local probabilistic evaluation of other network nodes known to him. The probabilistic assessment gives an estimate of the probability of delivering a message on condition it is forwarded to a particular node.

Probabilities of delivery are regularly updated. Attributes used to evaluate the best carrier (e.g. the node priced as the node with the highest probability of delivering the message) are implemented as frequency of connections with other nodes, the probability of occurrence in the range of the target area, the degree of mobility. In [133] the authors then show that the routing protocol CAR is more scalable than Epidemic Routing routing protocol. The protocol overhead CAR remains approximately the same, depending on the size of the local memory .

2.4.2.8 Hi-Bop

Another routing protocol, which is based on the principle of a context evaluation, is the routing protocol Hi-Bop [17], [20], [18], [19] designed for a special class of models, so-called systems of small worlds [98], [212]. Context information is divided into information about the current context node (the current context) and information on the evolution of the context node at a time (context evolution). Protocol Hi-Bop was designed for social networks. Each node is identified by a set of personal data (name, address, city, occupation, hobbies) that can be expanded with additional data. Each node keeps the information in the local data structure called the Identity Table (IT). When changing IT neighboring nodes, then the node obtains information about their neighbors and includes it in the internal representation of the current context. Current context provides information about the current network status and social interactions of the local user. These characteristics may vary over time, and therefore is kept as history. The description of the routing context can be found in [19].

2.4.2.9 BUBBLE-RAP

Hui et al. [76] proposed a routing scheme called BUBBLE-Rap, which combines two forwarding metrics: the node centrality and labels based on affiliation to community. Authors used two methods to detect communities: k-clique community detection algorithm proposed by Palla [146] for the detection of overlapping communities, and a modularity based approach proposed by Newman et.al [142]. BUBBLE-Rap routing protocol uses LABEL algorithm at the highest level to identify the node community and RANK routing algorithm to transmits messages. When two nodes encounter, the message is forwarded to nodes with higher centrality values than the current node. Authors reported that they analysed four real-world datasets (Cambridge, MIT Reality, Infocom, Honk-Kong) observed heterogeneity for node centralities in both global and local states. All four datasets are small, containing tens of unique nodes and were obtained for special kinds of communities:

conference participants or university students and staff. The authors proposed both centralised and distributed versions of routing scheme. The authors compared the correlation of centrality and the number of contacts and the centrality and the number of encounters and found that number of encounters correlated well with node centrality. For the computational reasons, the centrality has been supplemented by the number of encounters, which reflects the intuitive assumption, that does not depend how much people a member of OPN knows, but rather how frequently he interacts with these people.

2.4.2.10 MobySpace

Another routing protocol, which is based on the principle of context, is the routing protocol MobySpace Routing [102]. In this protocol, the information about possible contacts of pairs of nodes is kept as a representation of multidimensional Euclidean space called MobySpace. Two nodes, which have similar contacts are located in space of MobySpace together. The best node to receive the message is then the node which is closest to the destination node in MobySpace.

2.4.2.11 MobiClique

Pietiläinen et al.[157] proposed a multi-layer social routing scheme called MobiClique, which uses also the users' Facebook profiles consisting of a unique user identifier, the friend list and a list of groups of users sharing some common interests.

2.4.3 ML-SOR

ML-SOR is a routing algorithm proposed for routing in multi-layer social networks by Socievole et al. [186]. In real world in addition to face-to-face communication, users usually interact through other different channels: e-mail, phone calls, social network like Facebook, Twitter, LinkedIn etc. In order to describe these different autonomous communication systems and differentiate it from the standard face-to-face interaction, authors introduce terms detected social network and online social network. Detected social network denotes the node proximity graph containing communities and/or centralities computed on the social graph detected from the collected data on communicating mobile devices. Online social network denotes the graph constructed using social information extracted from virtual or self-declared contacts, such as a list of friends on facebook or followers on twitter. Detected social network and online social network represent different social contexts. Authors proposed a multi-layer network model to describe this connection of a single user to another users through different autonomous systems. The presented multi-layer social graph is based on the social graphs extracted from Bluetooth co- presence data, Facebook friend lists and shared interests. They presented two multi-layer models based on data from Lapland and Sigcomm datasets. Both multi-layer models consist of 4 layers: i) DSN Mode 1, ii) DSN Mode 2, iii) Facebook network, and iv) Interest network. Facebook network has been extracted using the participants' Facebook social network information. Each vertex represents a single user and an edge between two nodes represents the facebook friendship. Participants' interests to generate an Interest network social graph. In order to simplify the comparison between layers and to avoid modeling detected social network as a temporal graph, the authors have chosen to model a multi-layer social network using only static graphs. They have extracted Mode 1 and Mode 2 using Joint Diagonalization method proposed in [51]. They proposed a metric for each layer and join these metrics together into one ML-SOR social metric. The final ML-SOR routing metric is computed using a combination of three measures: centrality on DSN layer, tie strength on OSN layer and link predictor on Interest network layer.

2.4.3.1 Social Role Routing

Bigwood et al. [14] analyzed the structural similarity of networks and proposed opportunistic routing schema Social Role Routing [13].

They constructed two level-level model consisting of self-reported social network (SRSN) and detected social network (DSN). SRSN is constructed from self-declared facebook contacts, the DSN denotes the detected social network. Their approach is based on application of information from preexisting social networks to make routing decisions. They determine the social roles in the social network to find classes of nodes that may be useful for forwarding. They proposed and implemented a role connectivity graphs; a copy of this graph is stored in each node. Their approach is based on hypothesis that self-reported social networks are sufficient to bootstrap the opportunistic network.

2.4.3.2 SRAMSW

Guan et al. [64] proposed a spray-based routing algorithm by combining feedback information and retransmission timeout with buffer management mechanism and social-based metrics. The proposed routing method SRAMSW uses ack messages in order to detect timeout to solve the blind spot problem and proposes buffer management to reduce the overload by adaptively adjusting the message copies' lifetimes. Furthermore, it uses the information on social relationship in order to optimize the forwarding performance. Authors have compared the proposed routing method to some benchmark routing algorithms such as Epidemic Routing, PRoPHET, and traditional Spray and Wait and reported better performance than the traditional spray routing algorithm and the Epidemic routing algorithm on message delivery rate.

2.4.3.3 Explore and Wait

Borrego et al. [22] proposed Explore and wait routing scheme, which combines profile-cast model addressing with probabilistic routing metrics similar to those one use for routing in PROPHET and optimal stopping theory-based delivery strategy. As opposed to PROPHET, the proposed probabilistic scheme uses message forwarding instead of message copying. Only one copy of the Real-cast message is kept. Profile-cast model allows message destination be users or the groups of users defined by their profiles. The authors tested Explore and Wait routing scheme on two types of simulation scenarios: Urban OPN and Rural OPN and compared it to three simple forwarding strategies, including first contact or CSI routing scheme.

2.5 Human Mobility Models

This section describes different categories of mobility models.

2.5.1 Synthetic Models

Synthetic mobility models are mathematical models of mobility, which are created without the use of real-world data. Fiore et al. [53] propose the taxonomy of synthetic mobility models. They classify synthetic mobility models into five groups: i) stochastic models, ii) traffic stream models, iii) front models, iv) queue models and v) behavioral models of mobility. *Stochastic models* are the mathematical models that are built on the concept of a random movement of nodes. The examples of stochastic models is Random Walk (RWM) or Random Waypoint (RWP) mobility models. *Traffic stream models* describe the mobility of vehicles using mathematical and physical apparatus proposed by the analysis of streaming in hydrodynamics. *Following car models* include models where together with modeling the movement of the vehicle the behavior of the driver is modeled. *Queue models* of mobility model roads as FIFO queues and vehicles as clients. In *Behavioral social models* the movement of each node is fully or partially determined by the synthesized behavioral rules or social influences.

2.5.1.1 RWM: Random Walk an its Extensions

The random walk model [57] is the simplest model of mobility. This model generates a completely random node movement patterns. The node mobility simulation based on RWM starts with random initial distribution of nodes in the simulation space. In each step of simulation, each node moves in random direction with the randomly selected speed. The major drawback of this approach is that the RWM model generates node mobility patterns that do not correspond to the real human movement trajectories. Also sharp and sudden changes of direction are far away from the real-world human mobility patterns. [149]. Nain, P. et al. [136] proposed two RWM extensions: Random Walk with Wrapping and Random walk with Reflection. The advantage of the described model is its easy implementation. Similarly to RWM, the major drawback of this approach is that it generates node mobility patterns that do not correspond to the real human movement trajectories. Sharma et al. [182] proposed Hybrid Random Walk. This mobility model is based on one-dimensional parameterisation.

2.5.1.2 RWP: Random Waypoint and its Extensions

Sharma et al. [183] proposed a RWP mobility model with the time delays between node movements. Popularity of this model of a large number of its modifications. Here we mention some of them. One such modification is the tuning of key parameters of the RWP model on the basis of a real data [170] . In [73], the authors presented an extension of the RWP mobility model. Another modification of the RWP model is the model called *Swiss Flag*, which was introduced by Le Boudec in [99]. The proposed modification reflects the demand to suppress the effect of a “speed decay”, which is an inherent part of the classic RWP mobility model simulations. In the model proposed by Le Boudec, the simulation area was introduced as a region in the shape resembling a cross on the Swiss flag, from which is derived the name of the model. Another modification of the RWP mobility model, so called *Restricted Random Waypoint* (RRW) mobility model was proposed by Blazevic et al. [16]. This model is based on the assumption, that the nodes in the large areas do not move to the arbitrary destination points but tend to move towards destination points located in their neighborhood.

2.5.2 Survey-based Models

Survey-based Models are based on statistical observations. They describe observed mobility patterns by the set of statistical or probabilistic functions. These models are to generate pseudo-random or deterministic behavior of nodes.Complex macroscopic mobility model was introduced in order to model vehicular traffic in a real network of roads and highways in Switzerland. This model was developed at ETH Zurich. The model is continuously calibrated using real statistical data obtained by monitoring real traffic. A similar approach uses a model developed by Los Alamos Research Labs. The second one uses a more accurate statistical data obtained from data taken by the sensors at the traffic lights at intersections and data from automated systems for the evaluation of traffic density. Statistical analysis of trajectories Bazzani et al. [10] presented statistical analysis of a mobility dataset obtained in the Florence urban area. They tested by the probability distribution and the moving object activity of robust statistical laws. Due to the rapid development of wireless technologies and high-tech applications, many publications have appeared in recent years, which are related to analysis of large datasets. Liu et al. [116] analyzed data consisting of 85 million GPS points of taxicabs collected in Wuhan, China. They proposed mobility model based on spatio-temporal paths of moving nodes and spatio-temporal clustering algorithm, which uses spatial clustering of node positions at different times and a method of complex hull to merge these clusters into spatio-temporal ones. Hoque et al.[71] analyzed GPS data of taxicabs obtained in the San Francisco area by application of clustering and statistical methods. Cheng et Anbaroglu et al. [40] proposed a spatio-temporal clustering algorithm for complex temporal networks analysis in spatial, temporal and thematic domains and tested it on data obtained from a part of London’s traffic network.

2.5.3 Activity Based Models

In recent years, particular attention to human mobility analysis in urban areas has been paid in urban transportation modeling in order to improve transportation planning. Several publications have appeared in recent years documenting that human movements are organized based on activities and locations that are important in their daily life. Transportation activity-based models consider travel demands as the needs derived from human activities. The increasing popularity mobility models for transportation planning are based on :motifs". The concept of network motif was firstly proposed in complex system research by Milo et al. [125]. Network motifs are defined as recurrent and statistically significant sub-graphs or patterns. The concept of network motifs has been adopted and widely used in the analysis of complex biological networks [178], [43], [87], [213], [155], [176]. Schneider et al.[175] proposed the application of network motifs in human mobility analysis. They constructed daily human mobility networks from CDR data for Paris over a period of 6 months and from travel survey data for Paris for one day. They reported, that they identified 17 unique motifs. Jiang et al. [84] have applied a similar approach to extract human daily motifs. They have constructed daily human mobility networks from triangulated mobile phone CDR data for one million users in Boston. They have reported similar findings. Furthermore, they have proposed a probabilistic inference method to use motifs, time of day, activity sequence, and land use related information to further infer activity types and traffic patterns Widhalm et al. [214] proposed methods for inferring human activity types from data extracted mobile phone data and land use data for the cities of Boston and Vienna.

2.5.4 Trajectory based models

Trajectory models, also called *Trace Based Models* use the extraction of mobility patterns from the set of paths. The major drawback models lies in the fact, that the extracted mobility models are valid only for a narrow group of nodes, which has the properties corresponding to the properties of nodes whose routes were used as input data for extrapolation. Research work on this subject can be divided into two large groups. The first group includes the work include large-scale projects focused on of monitoring traces of vehicles [204], [225] and the similar large-scale projects focused on humans as MITE Reality Mining [128], USC MobiLib [69] or Cabspotting [70]. The second group include projects, which are primary targeted to human mobility models extraction [96], [100], [34], [195]. The research is focused on the analysis of experimental datasets containing data, which were collected on small groups of people. Participants of these experiments received instructions to carry the wireless nodes such as mobile phones or sensor devices in order to detect the nodes in proximity range and log the data on these detected nodes during the experiment duration. Usually these datasets were collected in university communities and the participants of experiment were both students and teachers. These datasets have been used in OPN research, we shortly describe some of these datasets in order to help a reader have a better orientation in the profile of typical datasets. University of Milano Campus (also referred as UNIMI) dataset was collected for 19 days in November 2008. The dataset contains data from 44 people. The connectivity range of communication devices was about 10 meters. The carriers of the devices were faculty members, doctoral students, and technical staff. University of Cambridge Campus dataset (also referred as Cambridge) contains data from 12 people. Dataset was collected for 6 days. The communication devices have Bluetooth connectivity with 10 meter contact range. University of St Andrews Campus dataset (also referred as SASSY) contains data from 27 people (22 undergraduate students, 3 graduate students and 2 staff) at University of St Andrews. It also contains the data on facebook friendship of the participants. Marseilles High School dynamics contact network dataset (also referred as Marseilles) contains the data collected from students in a high school in Marseilles, France. It consists of two separated datasets. The first dataset was collected among the students of three classes during 4 days in 2011. The second dataset was collected for 7 days in 2012 and contains data from students of 5 classes. UNICAL [150] dataset contains data from 15 postgraduate students at University of Calabria campus, Italy. It was collected for several days in 2014. The dataset also contains the data on facebook friendship of the participants. SIGCOMM dataset [156] contains data from 76 conference attendees collected by an opportunistic mobile phone social application. In the last decade, the approaches to trace extraction based on semantic description [188] clustering [168], [11] or spatio temporal clustering [97] has attracted much attention. Yan et al. [222] proposed hybrid model of trajectory and were interested in spatio-temporal patterns. There are two important results of the analysis of the real-world mobility data. The first result include knowledge that both the movement speed of the nodes and the lengths of breaks of the nodes have the *log-normal distribution*. In the synthetic models proposed before the real human traces were analyzed, the uniform distribution was assumed. Another important finding is that the time for which the network nodes remain in contact, can be modeled by means of the probability distribution of the type of *power law*, and not by the exponential distribution. These findings have had a retroactive effect on the synthetic models and their configuration parameters.

2.5.5 Traffic Simulators

Traffic Simulators model individual entities participating in the simulation. For the modeling of transport in urban areas were developed many models such as Paramics, Corsim, VISSIM or TRANSIMS [53]. In recent years, commercial traffic simulators have been developed as software applications designed primarily for traffic analysis and planning. They are used primarily as a tool for decision support and approval of plans for the construction of transport infrastructure. Some examples of noncommercial traffic simulators include MobSim introduced in [132] or the ONE simulator [94].

2.5.6 Mobility Models Constructed by Machine Learning Techniques

Toch et al.[200] proposed a taxonomy of human mobility models constructed by the application of machine learning technology. At the highest level,the models are divided into three categories: i) user modeling, ii) place modeling and iii) trajectory modeling. Each category is divided into three categories: i) non machine-learning models, ii) supervised learning, iii) unsupervised learning and latent variable models. Each subcategory is divided into subcategories in accordance to machine learning technique in use. Authors cite a huge number of scientific papers covering almost all machine learning approaches to user, place or trajectory modeling.

2.5.7 Mobility Models from the viewpoint Of OPN

Hong et al. [219] has introduced five characteristics describing movement of a set of moving nodes forming OPN from the viewpoint of routing in OPNs. These characteristics include: i) Flight length, ii) Spatial locality distribution, iii) Temporal characteristics, iv)Spatio-temporal characteristics (these ones are called *Joint spatial and temporal characteristics* in Hong's terminology; we use notation spatio-temporal, which is more common for the data analysis both in spatial and time domain simultaneously), v) Graph characteristics.

The *Flight length* is defined as the longest straight line trip from one location to another.

The *Spatial locality distributions* describes the node scattering in scenario, which can be either uniform or heterogeneous. If every node has the equal chance to visit each location in the network, the spatial locality distribution is uniform, otherwise it is heterogeneous. Most real-world scenarios can be described using heterogeneous spatial locality distribution. Heterogeneous spatial locality distribution allow an application of clustering methods.

The *Temporal characteristics* describe temporal features of the scenario. The most important temporal characteristics include encounter frequency, filling time and scattering time

The *Spatio-temporal characteristics* are characteristics obtained by the mobility data analysis both in time and spatial domain simultaneously. Hong et al. have referred to spatio temporal analysis of an OPN employing the static wireless devices at certain location, so called throw-boxes, which can help message dissemination among mobile nodes in delay tolerant networks.

The *Graph-related characteristics* Graph related characteristics are used to analyze the potential presence of social networks in OPNs. In the context of OPNs, the social network denotes a set of nodes which are close to each other in some way. Typically, the nodes form a social network, when they visit common places, or they move close to each other or have meet each other frequently. The existence of a social network facilitates application of metrics based on graph structural properties, which cannot be ignored since they can make a strong impact on message delivery. Hong et al. [219] has enumerated three main groups of graph-related characteristics: centrality features, community detection, particularly k-clique communities and features based on the connectivity of a network graph formed by mobile nodes in a continuum framework [39]. Centrality is used to measure the importance of node in terms of the network structure. Hong et al. [219] has enumerated three common metrics: degree centrality, closeness centrality, and betweenness centrality.

Chapter 3

Overview of Our Approach

3.1 Introduction

The opportunistic networks (OPN) are networks disseminating messages with the "store-carry-forward" routing principle. The key function of OPN routing protocols is to make decisions on message forwarding. The routing metrics are designed in order to select the most optimal nodes which have the highest probability to be a part of the paths of successfully delivered messages with respect to maximization of message delivery ratio and the minimization of overhead cost and message delivery delay.

We proposed four routing schemes in this chapter:

i) Hierarchical Routing with Clustering 1 (HRC1),

ii) Hierarchical Routing with Clustering 2 (HRC2),

iii) SVM-based routing,

iv) Routing scheme combining GMRF (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm).

This chapter is organized as follows: at first, the short overview of intuitive constraints set to modeled OPN is given. Then the definition of OPN contact graph and the geographic are of OPN are given. Then, we discuss existing routing methods, which are related to our research because of they use in some way the context or regular patterns in node mobility. Next section describes two proposed methods Hierarchical Routing with Clustering 1 (HRC1) and Hierarchical Routing with Clustering 2 (HRC2). Then the inference of SVM-based routing method is presented. Finally, we describe how to apply GMRF to cell message delivery probability matrices and model these matrices as textures. Then the ANMA method is proposed.

3.2 Assumptions

As it is apparent from the text above, the definition of opportunistic network covers a wide range of networks from sensor networks for monitoring wildlife to the networks formed from autonomous robots. These networks differ in their properties.

In order to propose methods and procedures that are more specific, we proposed certain assumptions to define the research area. In our work, we adopt the following assumptions and constrains:

Assumption 1: Node character limitations
We limit our research to the opportunistic networks where nodes are:

- people with communication equipment allowing to operate as the node of the opportunistic network,
- autonomous ground mobile robots with the communication equipment allowing to operate as the node of the opportunistic network,
- vehicles driven by a driver with the communication equipment allowing to operate as the node of the opportunistic network.

Assumption 2: The autonomy of nodes
From the viewpoint of interaction with the environment and with the other nodes, the nodes must behave as autonomous devices or autonomous individuals. Each node must be able to make decisions on routes and traces of its movement in the environment. Each node in the network must be able to decide whether or not to communicate with other nodes within its communication range. Each node must be capable to make the decisions about communication: whether to initiate communication or whether to respond to the request for communication, or whether accept or not to accept a particular message. This assumption does not exclude the use of a fixed central data warehouse and operations over large data performed externally.

Assumption 3: Network where the certain patterns of movement are observable
We assume that the motion of the nodes is not random, but that the nodes move in order to undertake some reasonable activity. The movement of each node is determined by the sequence of node specific local destinations. It is assumed that each node is moved in order to meet certain specific task. Network nodes do not therefore move randomly, but they move along the meaningful trajectories; the source and destination points of these trajectories are determined by the task of the node. This assumption is based on behavioral models. The research focused on the movement of people and vehicles in the real world, indicates that persons and vehicles repeatedly occur in the same places, and we can observe certain patterns of movement.

Our research will be strictly limited to the opportunistic networks, where we can assume the existence of patterns of movement.

Assumption 4: Testing and development on the simulated data
Although the proposed algorithms are intended to be applied in the real opportunistic networks, they were developed and tested on the simulated data. Due to financial reasons, we are not able to test the algorithms in a real life.

Assumption 5: The physical dimensions of a network are not perceived as a limiting factor
We assume that the same routing principles can be applied both in micro-networks (networks of robots within a building) in medium networks, and macro-networks (networks in geographic areas). This assumption does not exclude that the requirements for hardware can be significantly different for different types of networks.

Assumption 6: The properties of the communication device
Each communication device representing the node in the opportunistic network has a unique hardware identifier. Individual devices can be uniquely identified. The communication device is capable:

- to exchange messages with other devices of the same type, if the device is in communication range,
- to keep the messages and data in a local memory,
- to determine the node position and store the information on the previous node position,
- to have a computational capacity adequate for the proposed routing methods.

Furthermore, the communication device could have the other properties, which follow. The device can:

- to communicate with some external communication devices out of the opportunistic network,
- to collect and utilize other parameters (such as user activity, calls, memory availability, state of the energy resources).

3.3 OPN: The Opportunistic Network

The opportunistic network OPN is the communication network where the messages are disseminated with the "store-carry-forward" routing principle. The OPN can be described by the temporal contact graph of nodes and by the multidimensional matrix of node coordinates in time, which represent node positions in the real-world physical area. If additional properties of nodes are taking into account, the model of OPN can be extended to the multi-layer model, where each layer represents another kind of relations among the nodes. The example is social aware routing in OPNs, which takes into account also other additional relations and connections among the nodes, such as friendship, interests or e-mail communication.

The routing methods proposed in this work are related to the OPNs, where only data on node mobility are available as input data. The practical application scenarios of these kind of OPN include traffic condition monitoring, advertisement, environment condition monitoring, newscast, place-driven assistant systems. No assumption is made about additional connections of people who serve as the nodes and register in OPN in order to use the services of OPN. The main routing criterion is co-locality of nodes. The node communities formed by co-location are taking into account.

OPN contact graph

The OPN contact graph $G(N, E, F(t))$ is the temporal graph, where $N = n_1, n_2, ..., n_j$ is the set of j opportunistic network graph vertices, while $E = e_1, e_2, .., e_k$ is the set of k graph edges and $F(t)$ is a set of edge functions $(f1(t), f2(t) \dots)$ which defines the existence of the edge at time t. The vertices correspond to the nodes of the opportunistic network.

Geographical area

The geographical area of the opportunistic network Λ is the area in which the opportunistic communication takes place. It is represented as a finite set of ordered pairs m, n. Nodes can move in this area, leave it, enter it or stop inside the area and temporarily be inactive.

Traces

Nodes move along the traces. Trace r is a set of points from the area of the opportunistic network through which the node moved from m_1, n_1 to point m_2, n_2. m_1, n_1 is a start point of the trace. m_2, n_2 is an end point of the trace r. Group of traces forms a trip.

In the recent years, the particular attention to the analysis of communities within networks was given in various disciplines, particularly but not only in mathematics, physics and biology. Scientist have become interested in the study of networks describing topologies of wide variety real systems [52]: biochemical networks, social networks, communication networks, transportation networks, text databases networks, world wide we and much more.

3.4 Existing Methods

Similarly to existing more sophisticated OPN routing schemes, the basic idea of the proposed methods consists in routing efficiency improvement through forwarding messages only to those nodes of the OPN, which fulfill some criterion (or criteria) constructed from the data. The literature on OPN routing shows a variety of approaches; we are interested in those ones, which use in some way context or prediction or node mobility models. From the viewpoint of the main principle of routing metrics in use, the context-aware OPN routing techniques include into seven groups:

i) Probability-based routing protocols

ii) Network community oriented methods

iii) Node influence oriented methods

iv) Multi-layer social aware oriented approaches

v) Geographic routing methods

3.4.0.1 Probability-based Routing Protocols

The probability based routing protocols use metrics based on history, but they avoid construction of mobility patterns at all or use simple prediction models derived from the Markov process model. The examples of probability-based routing protocols are PROPHET, MobySpace, LIBRE, MaxProp or PER. PROPHET and MaxProp improve routing using routing metric based on probabilities of node contacts computed from the node contacts in history. MobySpace routing algorithm improves routing using routing metric based on probability of visiting localities computed from the locality visits in node history. LIBRE explores contact probabilities computed from node encounters in history and use Markov model for node mobility prediction.

3.4.0.2 Node Influence Oriented Methods

The centrality is a network property which characterizes the node importance in the network. The most recognized centralities are closeness centrality, degree centrality and betweenness centrality.

The *closeness centrality* of a vertex v, for a given graph $G := (V, E)$ with N vertices and $|E|$ edges, is defined as

$$C(x) = \frac{1}{\sum_y d(y,x)} \tag{3.1}$$

where $d(y, x)$ is the distance between vertices x and y.

It is impossible to compute *closeness centrality* for disconnected graphs, because the distance between two nodes, which belong to two distinct components of graph, has not a finite value.

$$H(x) = \sum_{y \neq x} \frac{1}{d(y,x)} \tag{3.2}$$

where $1/d(y, x) = 0$ if there is no path from y to x.

Betweenness is a centrality measure of a vertex within a graph, which quantifies the number of how many times the node acts as a bridge along the shortest path between two other nodes.

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \tag{3.3}$$

where σ_{st} is total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v.

Both betweenness and closeness centralities of all vertices in a graph involve calculating the shortest paths between all pairs of vertices on a graph. The nodes which have high values of betweenness centrality become to the most important nodes for routing in OPNs. The equations above are valid for static graphs; for temporal graphs it is necessary to compute centralities through the time development of the graph. It is computational expensive. Using centralities as routing metrics give very good results and it is involved in several routing schemes.

Tang [197] extended the definition of closeness to temporal graphs using the temporal shortest path length between nodes, which is a measure of how fast a source node can deliver a message to all the other nodes of the network. The disadvantage of application particularly temporal betweenness centrality is increasing computational time for large temporal networks.

3.4.1 Community-based Approaches

The main idea of community-based routing is that relationship of the users is reasonable information for predicting future contact opportunities. The community has a strong impact on human mobility pattern. The community-based routing schemes consists of two phases. At first, the mobile nodes are grouped into communities by certain community detection algorithm. Secondly, the routing scheme is proposed and the messages are forwarded in accordance to this routing schema. In the recent years, the particular attention to the analysis of communities within networks was given in various disciplines, particularly but not only in mathematics, physics and biology. Scientist have become interested in the study of networks describing topologies of wide variety real systems [47]. Biochemical networks, social networks, communication networks, transportation networks, text databases networks, world wide we and much more.

More real networks typically contain parts in which the nodes are more highly connected to each other than to the rest of the network. These parts of networks are represented by subsets of nodes, which are called are called clusters, communities, cohesive groups or modules [142]. In accordance to Palla et al. [146], no unique definition is accepted, different terminology is used in different scientific areas. In this work, we will use the term community, because this term is mostly used in scientific community interested in opportunistic networking.

Both Danon [47] and Newman [142] presented a survey of several existing methods. Generally, the problem of finding non-overlapping communities in a network is a graph partitioning problem. It consists in the task of partitioning the given graph into n sub-graphs. Graph partitioning problem has been studied in graph theory, computer science. Similar problem was has been studied in sociology, where it is called hierarchical clustering or hierarchical communities detection. Working with data collected on real systems, usually the following problems bust be taken into account:i) we usually don't know, how many communities should be discovered, ii) the communities can be organized hierarchically iii) computational complexity of community detection method. Palla et al. [146] proposed k-CLIQUE algorithm for static graphs, where another approach finding overlapping communities where a community is defined as the union of all k-cliques (complete sub-graphs with k nodes) that can reach each other through a series of adjacent k-cliques, where two k-cliques are said to be adjacent if they share k-1 nodes.

Newman et al. [143] proposed a method based on modularity, which is able to find communities without the necessity to specify the number of communities. The method is based on edge removal from the graph. The edges to be removed are selected on the basis of evaluation using betweenness coefficients. The computational complexity of the proposed method is $O(m^2n)$.

3.4.2 Multi-layer Social Aware Oriented Approaches

Multi-layer approaches to routing in OPN model the OPN as a structure of mutually connected layers. In addition to contact graphs describing physical encounters of nodes, they use social layers reflecting the real world contacts. The presented multi-layer social graph is based on the social graphs extracted from different sources such as Facebook, Twitter or e-mail communication. The examples are ML-SOL, Social Role Routing or MobiClique.

3.4.3 Geographical Routing

In opposite to topology-based routing schemes, the geographic routing protocols use the knowledge of physical location in the routing process. Geographic routing protocols are context-aware routing schemes, which use location data as the context. One group of routing schemes use location services. The examples of routing protocols based on location service include Location services some specialized nodes, which serve as location Li et al. [109] a scalable location service for geographic ad hoc routing, which allows determine the location of the network destination. Basagni et al.[8] proposed DREAM a routing scheme based on flooding, which uses partitional flooding. the flooding is restricted to some some region. Leong et al.[107] proposed method for nodes, which has no information on their locations which is called virtual coordinates.

3.5 Hierarchical Routing with Clustering

3.5.1 Overview of the Method

Problem:

i) to propose application of unsupervised machine learning in routing metrics inference,
ii) to analyze the performance of the proposed routing method.

We proposed two hierarchical routing methods *Hierarchical Routing with Clustering 1 (HRC1)* and *Hierarchical Routing with Clustering 2 (HRC2).* The block diagram of the proposed method HRC1 is presented on Fig. 3.1. Fig. 3.2 shows the second method HRC2, which is an adaptation of HRC1.

3.5.1.1 Simulation Setup and Dataset

Dataset obtained from node movement simulation. After the simulation scenario is prepared, we run the simulation in the simulation environment ONE [94]. The output of simulation is a file containing records of node positions at each time step. The records have the following form:

$t_i, x_1, y_1, x_2, y_2, x_3, y_3 \dots\dots\dots x_N, y_N$

where t_i is a time step, x_j, y_j are the Cartesian coordinates of the node X_{ij} and N is the number of nodes. This records represent a dataset containing temporal data on node positions.

3.5.1.2 Contact Graph

We compute the contact graph of nodes at each time step. The vertices represents the nodes, the edges represent contacts. If two nodes A and B are in communication distance at time t, the contact graph for time t contains an edge between A and B.

3.5.1.3 Cluster Analysis

Cluster analysis is the task of grouping a set of objects in such a way that objects which are more similar to each other are assigned in the same group. These groups are called clusters.

Given a set of observations $(y_1, y_2, \dots, y_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets $(k \leq n)$ $S = \{S_1, S_2, \dots, S_k\}$ so as to minimize the within-cluster sum of squares

$$\arg_S \min \sum_{i=1}^{k} \sum_{y_i \in S_i} \parallel y_j - \mu_i \parallel^2 \tag{3.4}$$

where μ_i is the mean of points in S_i.

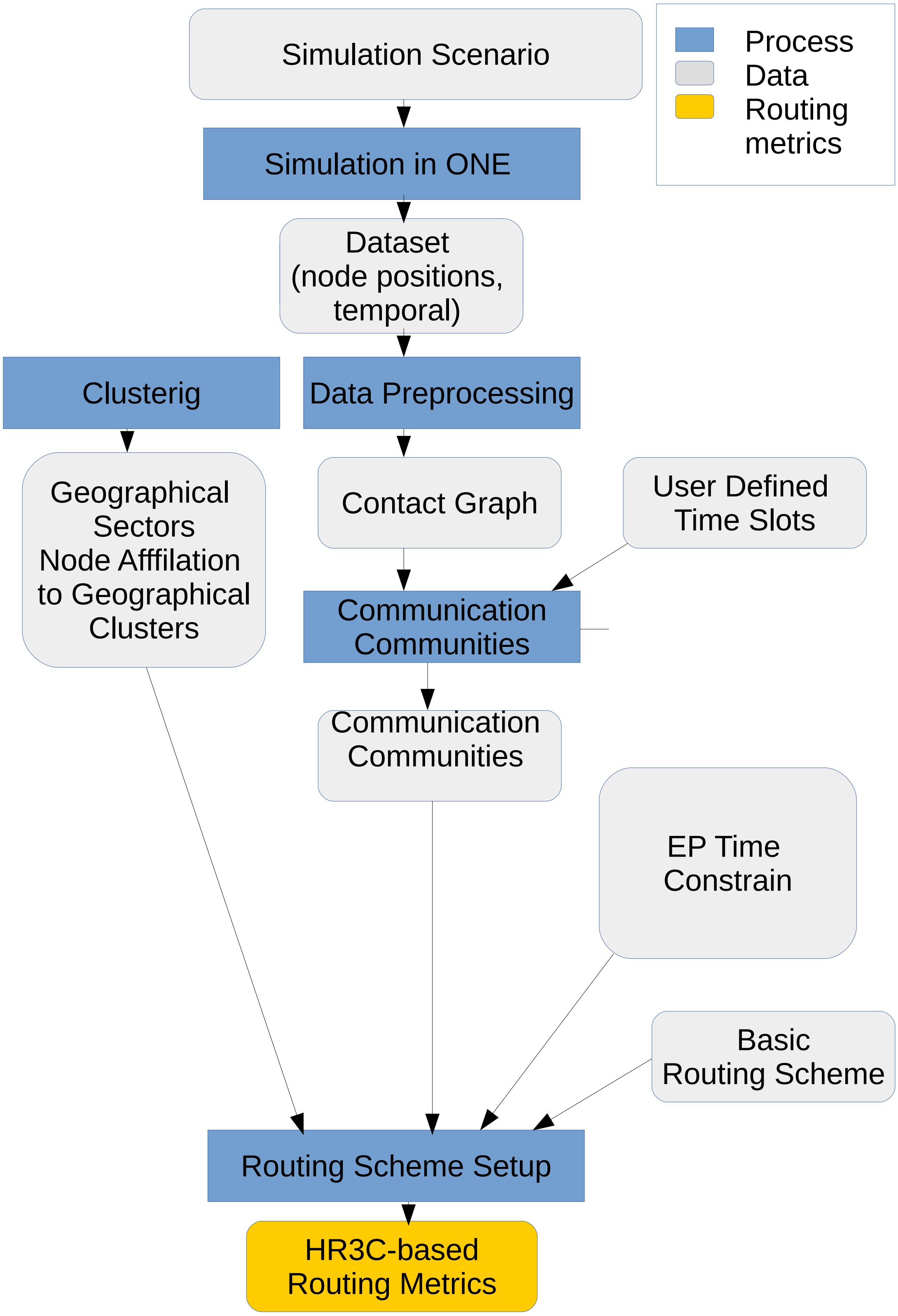

Figure 3.1: The block diagram of the process of routing metrics HRC1 inference

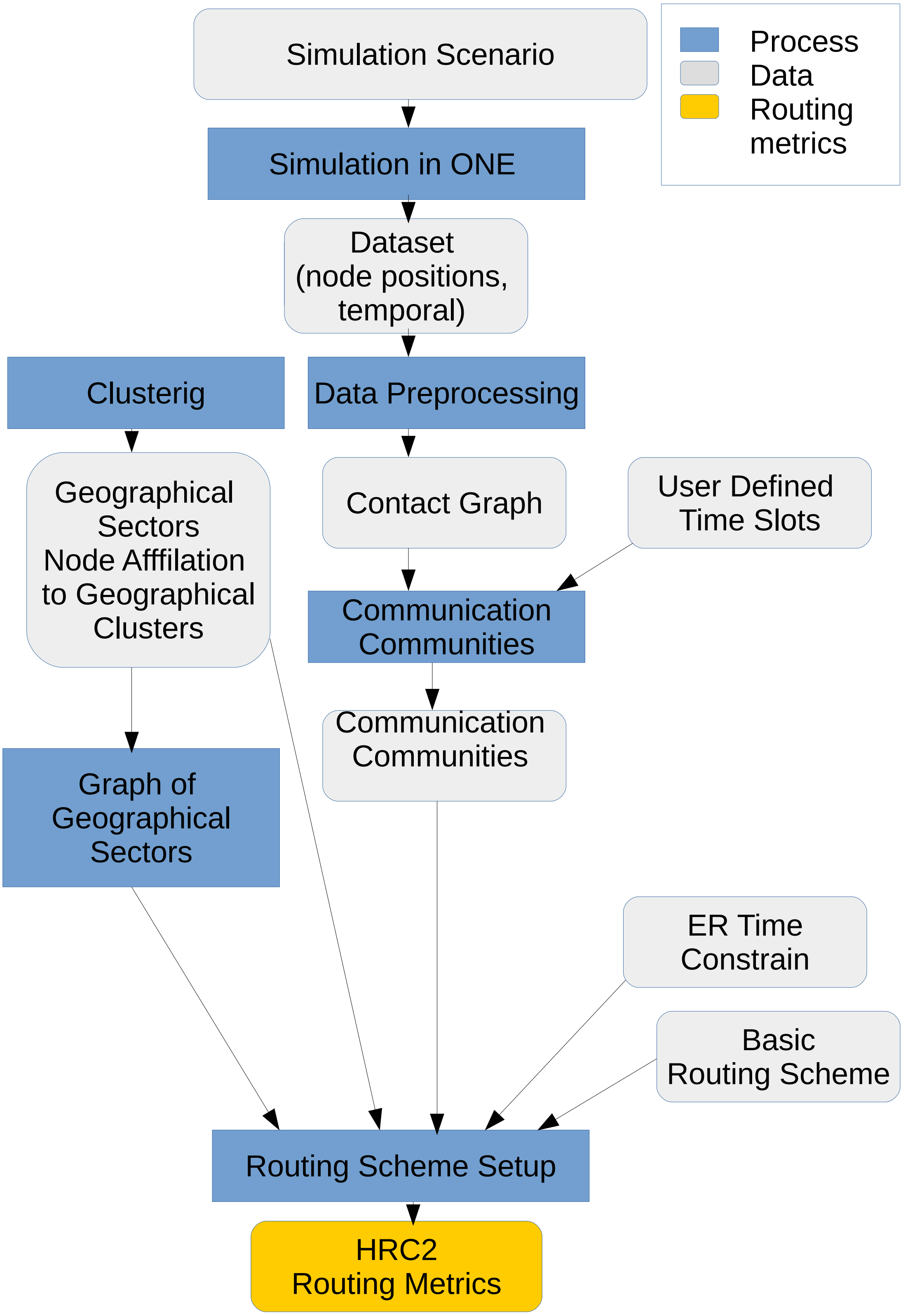

Figure 3.2: The block diagram of the process of routing metrics HRC2 inference

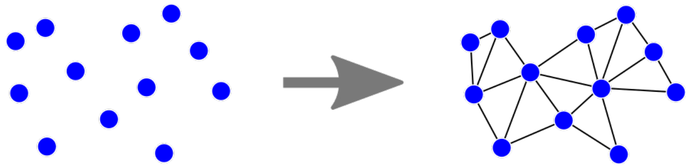

Figure 3.3: Detection of OPN Geographical Sectors: Data Pre-Processing

There are many algorithms for cluster analysis: in accordance to [218] possibly over 100 clustering algorithms has been published. k-means algorithm [67] uses an iterative refinement technique. Given an initial set of k means $m_1(1), \dots, m_k(1)$, the algorithm proceeds by alternating between two steps: the assignment step and the update one.

In the assignment step, each observation (piece of the data) is assigned to the cluster with the closest mean.

$$S_i^{(t)} = \left\{ y_j : \| y_j - m_i^{(t)} \| \leq \| y_j - m_{i^*}^{(t)} \| \right\}$$
$$\text{for all } i^* = 1, \dots, k \tag{3.5}$$

In the update step, the new means are calculated to be the centroid of the observations in the cluster.

$$m_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{y_j \in S_i^{(t)}} y_j \tag{3.6}$$

3.5.2 Detection of OPN Geographical Sectors

This section describes partitioning of the OPN geographical area in accordance to its geographical structure. Our approach is based on application of unsupervised machine learning to node positions data. We used well-known k-means algorithm with Euclidean distance. First, we perform the data preprocessing: we construct the node position graph from the geographical data. We proposed iterative triangulation to obtain a graphical representation of node positions. In node position graph, the vertexes represent node positions in time (x,y,t) obtained in different time steps. The edges represent connections between vertices obtained by iterative triangulation with the limited value of the length of the triangle side. Fig. 3.3 shows an example of a triangular graph of node positions collected in time.

The points inside the triangles are added to the data. It makes the following clustering more robust. In the next step, the geographical sectors of the OPN are computed over the spatio-temporal data using clustering, we use a variant of k-means method. Because of the spatio-temporal nature of the data, each node can appear in more than in one cluster (but with a different time coordinate). Thirdly, we compute a node affiliation into detected geographic sectors. In order to eliminate random or rare node presence in sectors, we propose a metric of node affiliation into cluster: the number of node positions inside geographical sector is divided by all the node positions in time. If this value is higher than δ, the node is affiliate with the sector i. It is necessary set the value of δ manually, we use the 0.05, e.g. at least five percent of all node positions must be inside the geographical sector. This value is dependent both on OPN node mobility and geographical high level structure and must be set with respect to these characteristics. For each node X we compute GSEC(X), a set of labels of geographical sectors which the node is affiliate to. Finally, we constructed a graph of the connection of clusters. The graph nodes represent clusters. The graph edges represent connection of clusters by X set of vehicles moving in an opportunistic network. In comparison to social-aware routing protocols such as BUBBLE RAP or ML-SOL, which analyze communities observed in a contact graph constructed from the node contact data, we analyze the communities in the graph based on geographical closeness of node positions without any regard to social contacts except those ones described by repeating co-locations of the nodes. It is sufficient, because we are interested in geographical partitioning of the OPN in this step.

3.5.3 Construction of Communication Community in spatio-temporal domain with time constraints

This section describes how the local communication communities are constructed from the spatio-temporal data. In opposition to method based on k-clique communities, where the communities are constructed from the k-cliques (fully connected sub-graphs) or to method based on modularity, we explore another approach based on communication contact graph construction in pre-defined time slots. In order to find communication communities, we investigated clustering method which uses the neighborhood node distance metrics, iterative node labeling and time constraints. The proposed method is based on building of community using community labels. We divide training data into time slots. The duration of each time slot is 15 minutes and it is divided into time windows of duration 2 seconds. At the beginning, each single node represents a community. For each node, the noes in communication distance in time window are found and added to the same community (relabeling). The computation is repeated for the next time window. As the process proceeds, the number of communities in node population decreases.

The iterative aspect of our approach is similar to the iterative graph partitioning method called Label Propagation Algorithm (LPA) proposed by Raghavan et al. [163]. In opposite to Raghavan's approach, our method does not assign the vertex the label that is most prevalent in the vertex' neighborhood, but rather is looking for connected node communities and operates in spatio-temporal domain. We call the constructed sub-graph local communication community. Notice, that found communities are not generally equal to those found by k-clique or modularity based community detection methods. The pseudocode of Find_Local_Communication_Community follows. The proposed method does not reflect the direction of time.

COMPUTATION OF LOCAL COMMUNICATION COMMUNITIES O(X)

- PSEUDOCODE

```
1. for ∀u ∈ U do
      o[u] = 0
   end
   ok = true;
   do
       for ∀u ∈ U do
           if o[u] == 0 then
               area = 0;
               for ∀v ∈ U do
                   if u ≠ v then
                       if distance(u, v) < limit then
                           if o[v] ≠ 0 then
                               if o[v] ≠area then
                                   if area≠ 0 then
                                       for ∀x ∈ o[area] do
                                           o[x] = o[v];
                                       end
                                   end if
                                   area = o[v];
                               end if
                           else
                               if area == 0 then
                                   area = generateNewNumber();
                               end if
                               o[u] = area;
                               o[v] = area;
                           end if
                       end if
                   end if
               end
               if o[u] == 0 then o[u] = generateNewNumber();
               end if
               ok = false;
           end if
       end
   while not ok;
```

For each time slot, we receive a Community Affilation matrix, where the columns represent the nodes, the rows time-slots and the values the affiliation of nodes to its communication community at each time slot. The labeling of communities differ for each time slot. Two same communities found in different time slots would have different labels. Fig. 3.4

shows several first rows and columns of Community Affilation matrix.

$$
\begin{array}{cc}
 & \begin{array}{ccccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & \dots & \text{time slot } \alpha \end{array} \\
\text{nodes}\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ \vdots \\ n \end{array} &
\left[\begin{array}{ccccccccccccccc}
8 & 8 & 8 & 8 & 4 & 4 & 4 & 8 & 8 & 8 & 8 & 4 & 4 & 4 & 8 \\
2 & 2 & 8 & 8 & 8 & 8 & 2 & 2 & 2 & 8 & 8 & 8 & 8 & 2 & 2 \\
2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \\
8 & 8 & 8 & 2 & 2 & 2 & 8 & 8 & 8 & 2 & 2 & 2 & 8 & 8 & 8 \;\dots \\
2 & 2 & 6 & 6 & 1 & 6 & 6 & 2 & 2 & 6 & 6 & 7 & 6 & 1 & 2 \\
1 & 1 & 1 & 5 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 7 & 5 & 0 & 0 \\
 & & & & & \vdots & & & & & & & & & \\
 & & & & & \vdots & & & \multicolumn{7}{c}{\text{local communication communities}}
\end{array}\right]
\end{array}
$$

Figure 3.4: Community Affilation Matrix

3.5.4 Clustering Node Positions into Locations

The measurement of node position co-ordinates with sampling frequency f generates a huge amount of data. Moreover, in real-world applications, if the node visits the same physical location again, the measured node position co-ordinates can gently vary from the node co-ordinates obtained from the previous measurements in the same location. In order to process data from more measurements to learn a mobility pattern, its better to recognize the presence of the node in location rather than work directly with exact node co-ordinates. For each node, we apply a variant of k-means clustering algorithm to find clusters of node positions and we call these clusters *locations*. We use the variant of k-means clustering method proposed by Ashbrook et al. [5] who have been interested in the task of finding locations from GPS data, which is some kind of locally applied k-means clustering. The k-means clustering is not applied to all node position data simultaneously: only the node positions in defined local neighborhood are clustered at each step. In image processing this variant of k-means clustering is sometimes called mean shift clustering. The original method contains an application of a circle neighborhood defined by the neighborhood center and radius, but we adopt a square neighborhood in our work, which is well-known from image processing, because it is more suitable for discrete data and can simplify 2D data processing in comparison to circle neighborhood. As result, we obtain a matrix TxN, where T is a number of where each row represents node locations in time. These row vectors are used to construct communication communities instead of original node positions vectors.

3.5.5 Routing Method

On the basis of data analysis described above, we proposed two hierarchical routing algorithms HRC1 and HRC2.

The routing method HRC1 combines three strategies in order to improve routing in OPNs: i) the node affiliation with detected OPN geographic sector + use of the sets of the detected geographic sectors, ii) the node affiliation with the communication community constructed in spatio-temporal domain with time constraints iii) epidemic routing.

The routing method HRC2 combines three strategies in order to improve routing in OPNs: i) the node affiliation with detected OPN geographic sector + use of the graph of the geographic sectors, ii) the node affiliation with the communication community constructed in spatio-temporal domain with time constraints iii) epidemic routing.

The proposed routing scheme has two phases: training and testing. During the testing phase, the knowledge is extracted from the data as it is described above. The labels and routing tables are computed centrally during the training phase. During the testing phase, the simulation is conducted. When two nodes encounter, they exchange messages in accordance to the rules defined by the proposed routing algorithm.

3.5.5.1 Message Exchange Policy

We proposed a message exchange policy, which is applied every time two nodes encounter. This policy enables to exchange a set of messages even if the message buffers of both encountering nodes are full using planning. If two node encounter, the lists of messages planned to be evaluated by the routing metrics are computed at first. The message exchange procedure is call makeDecision procedure, which includes a routing metrics. Both variants of HRC use the message exchange policy.

3.5.5.2 HRC1 Routing Method

FORWARD denotes message transmission, KEEP denotes that the node carrying a message does not forward the message. GSEC(X) denotes a set of labels describing an affiliation of node X to the particular geographical sector. $O(A)^{\alpha_A}$ denotes the local communication community of node X for the time slot α. We use time slots of length of 15 minutes.

Let‘s assume the nodes A and B encounter. They start communication and in accordance to the message exchange policy, each of them selects a set of messages to be exchanged. These messages are maintained in PLAN A list of node A and in PLAN B of node B list. Than for each message the routing rules are applied. At first, the nodes compare their labels of communication communities C(A), C(B), C(DEST). If all three nodes A,B, DEST are from the same communication community, the Local Encounter Metrics is applied. The DEST denotes the destination node of the message. The message is forwarded if the node popularity of the source node is smaller than the popularity of the node receiving message. The similar node popularity approach is used by LABEL or BUBBLE RAP routing protocols. If the nodes A,B and DEST are not from the same local communication community, the routing rule 2 is applied. Node A compares the affiliation GSEC(DEST) of DEST node to geosectors to the current meeting position. If the GSEC(DEST)= GSEC(current position), the routing scheme for searching local communication community is applied. Otherwise, the rule # is applied and the compare their affiliations to geographical sectors. If the B and DEST are affiliate to more common geographical sectors, the message is forwarded, otherwise the node A keeps the message.

MAKE DECISION PROCEDURE PSEUDOCODE

makeDecision (A, B, DEST)

1. **if** $C(A) == C(B)$ **and** $C(A) == C(DEST)$ **then**
return COPY & TIMEOUT;
end if

2. GSEC(DEST) contains the label of the geographical sector of the meeting point
Search the nearest time interval α_A where $O(A)^{\alpha_A} == O(DEST)^{\alpha_A}$.
Search the nearest time interval α_B where $O(B)^{\alpha_B} == O(DEST)^{\alpha_B}$.
if $\alpha_A == \infty$ **and** $\alpha_B == \infty$
then
if $O(A)^{\alpha} \cap O(DEST)^{\alpha+1}$ **or**
$O(A)^{\alpha} \cap O(O(DEST)^{\alpha+2})^{\alpha+1}$ **or**
$O(A)^{\alpha} \cap O(O(O(DEST)^{\alpha+3})^{\alpha+2})^{\alpha+1}$
then return FORWARD;
else return KEEP;
end if
else
if $\alpha_B == \alpha$ **or** $\alpha_B < \alpha_A$
then return FORWARD;
else return KEEP;
end if
end if

3. **if** $GSEC(A) \cap GSEC(DEST) \neq \emptyset$ **then**
if $|GSEC(A) \cap GSEC(DEST)| < |GSEC(B) \cap GSEC(DEST)|$
then return FORWARD;
else return KEEP;
end if
end if

3.5.5.3 HRC2 Routing Method

This section deals with the proposed hierarchical routing method HRC2. The block diagram of the proposed method is presented on Fig. 3.2

MAKE DECISION PROCEDURE PSEUDOCODE

makeDecision (A, B, DEST)

1. **if** $C(A) == C(B)$ **and** $C(A) == C(DEST)$ **then**
then return COPY & TIMEOUT;
end if

2. GSEC(DEST) contains the label of the geographical sector of the meeting point Search the nearest time interval α_A where $O(A)^{\alpha_A} == O(DEST)^{\alpha_A}$.

Search the nearest time interval α_B where $O(B)^{\alpha_B} == O(DEST)^{\alpha_B}$.
if $\alpha_A == \infty$ **and** $\alpha_B == \infty$
  then
    if $O(A)^\alpha \cap O(DEST)^{\alpha+1}$ **or**
      $O(A)^\alpha \cap O(O(DEST)^{\alpha+2})^{\alpha+1}$ **or**
      $O(A)^\alpha \cap O(O(O(DEST)^{\alpha+3})^{\alpha+2})^{\alpha+1}$
      then return FORWARD;
      else return KEEP;
    end if
  else
    if $\alpha_B == \alpha$ **or** $\alpha_B < \alpha_A$
      then return FORWARD;
      else return KEEP;
    end if
end if

3. if $GSEC(A) \cap GSEC(DEST) \neq \emptyset$ **then**
  if $|GSEC(A) \cap GSEC(DEST)| < |GSEC(B) \cap GSEC(DEST)|$
    then return FORWARD;
    else return KEEP;
  end if
end if

ROUTING ALGORITHM PSEUDOCODE

- start communication
- select messages to exchange - list PLAN A, PLAN B
- makeDecision(node U, node V, node DEST)
- makeDecision(node V, node U, node DEST)
- messageExchange
- end communication

The algorithm parameters follow:

- $U,\ V$ - nodes that meets.
- M - message identified by 32-bit ID.
- $M.dest$ - destination node for the message M.

- *U.clusters* - list of clusters visited by the node U during whole simulation.
- *PLAN* - list of messages from the buffer, which will be transmitted.
- list.*count()* - number of messages in the list.
- list.*size()* - maximum number of messages that can be stored in the list.
- list.*contains(M)* - test whether the message M is in the list.
- list.*insert(M)* - store message M into the list.

Routing Algorithm Pseudocode

1. Node U detected other node V (distance $|UV| < 0.128$).
PLAN = $\emptyset$;
2. foreach message M **from** U.buffer **do**
 if not V.buffer.contains(M) **then**
  if makeDecision(U, V, M.dest) == FORWARD **then**
   PLAN.insert(M);
  end if
 end if
end foreach
3. Limit the count of transmitted messages:
limitU = min(
 PLAN.count(), U.buffer.size() - U.buffer.count());
send(*limitU*);
receive(*limitV*);
limit = min(*limitU*, *limitV*);
if *limit* < PLAN.count() **then**
  delete last PLAN.count() - *limit* messages
   from the list PLAN;
end if
4. Send the first message M from the list PLAN;
Remove M from PLAN and from U.buffer;
5. if there is no transmission **then** go to 6;
else
  receive message M;
  if M.dest == U.id **then**
   message was successfully delivered;
  else
   decrement M.ttl;
   if M.ttl == 0 **then**
    ignore this message;
   else

if not U.buffer.full() **then**
U.buffer.insert(M);
else
ignore this message;
end if
end if
end if
end if
6. **if** PLAN.empty() **then**
if there is no transmission **then** go to 7;
else go to 5;
end if
else go to 4;
end if
7. End.

3.6 SVM-based Routing

3.6.1 Overview of the Method

Problem:

i) to propose application of supervised machine learning in routing metrics inference,
ii) to analyze the performance of the proposed routing method.

Let O be the OPN formed by the set of nodes $X = X_1, X_2, \ldots$ in geographical area Λ. The nodes are humans or vehicles. Previously it has been shown in scientific literature that humans do not move randomly, instead of that the human node mobility patterns show some regularities. Each node visits a limited set of places and moves in accordance to its own time schedule.

Fig. 3.5 shows node positions of a node X collected each day between 7 and 10 o'clock of 10 days of weekday simulation scenario. There are variances in node mobility data collected in different days, but some repeating patterns are observable.

Suppose we want to establish a communication channel between two geographical places A and B, $A \in K, B \in K$ through the opportunistic communication in the OPN O. Each time when any node X_i and place A are in one-hop communication distance, the node X_i and place A can exchange messages. The message from A addressed to B can be forwarded from A to X_i and than by routing in OPN to B.

Each time when any node X_j and place B are in one-hop communication distance, the node X_j and place B can exchange messages. If the node X_j carries the message addressed to B, it can forward this message to B. From the viewpoint of forwarding the message MSG_k through OPN, the communication between A and B is similar to opportunistic communication between X_i and X_j.

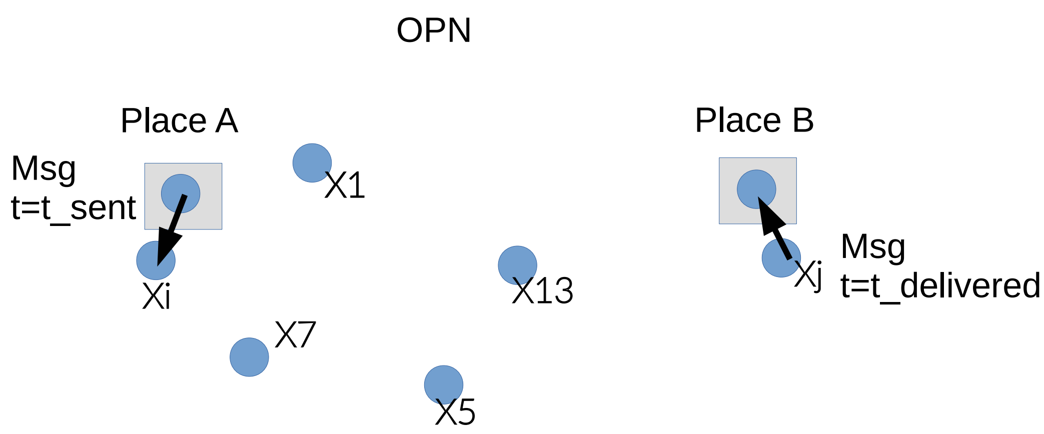

Figure 3.5: The visual representation of positions of one randomly selected node. The node positions were collected for 10 days between 7 and 10 o'clock of workday scenario (simulation)

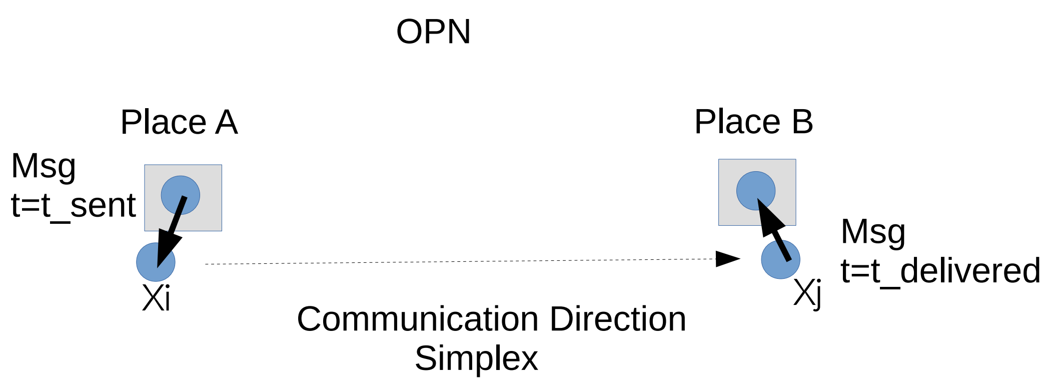

Figure 3.6: Stationary node A sends a message to stationary node B through OPN. From the viewpoint of forwarding the message MSG_k through OPN, the communication between A and B is similar to opportunistic communication between X_i and X_j.

We assume the existence of the node mobility patterns. This assumption implies that there should be a time pattern describing the appearance of the nodes in communication neighborhood of any geographical place during the day. If the messages are sent repeatedly from A to B during the simulation, it is possible to collect triples (X_i, Msg_id, time_sent) at place A and and triples (X_j, Msg_id, time_delivered) and try to estimate channel characteristics or predict future behavior of the channel. This channel is formed through OPN. By application an appropriate predictor or classifier, we can predict behaviour of OPN for the communication between A and B. Note, that the obtained data are dependent on the routing algorithm used for the dissemination of messages.

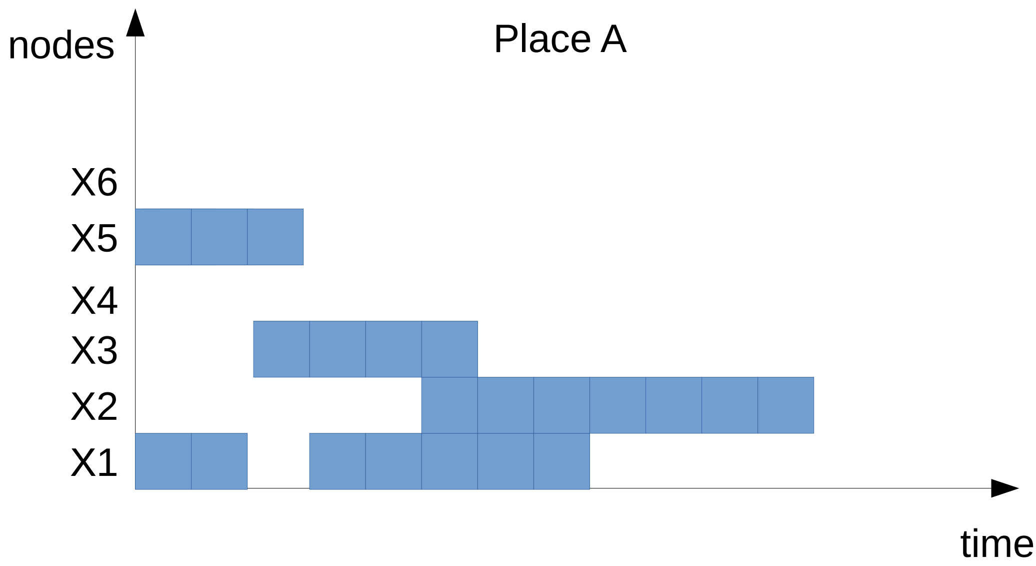

Figure 3.7: Block diagram of contact times of nodes X_i with place A

In order to improve OTN routing schema, we need implement a binary decision mechanism on message forwarding from A to X_i at time t. Suppose A intends to send a message MSG to B. If X_i meets the communication neighborhood of A at time t, the decision making mechanism must be able to make a decision about forwarding a message MSG from A to X_i. The task can be reformulated as a classification task: classify a triple (X_i, t, B), where X_i is the node, which receives a message from A, t is the time where the forwarding request occurs, and B is the destination place (extension to more destination places than just B) into two binary classes: 1 = MESSAGE DELIVERED, 0 = MESSAGE NOT DELIVERED. The classification scheme is shown on Fig. 3.8.

If there are no time constraints and the temporal graph of OPN is a connected graph, the classification is 1 for each input triple. We propose a time constraint *maximum message delivery delay m*. The training data is labeled with respect to maximum message delivery delay. The classification task is reformulated: classify a triple (X_i, t, B), where X_i is the node, which receives a message from A, t is the time where the forwarding request occurs, and B is the destination place (extension to more destination places than just B) into two

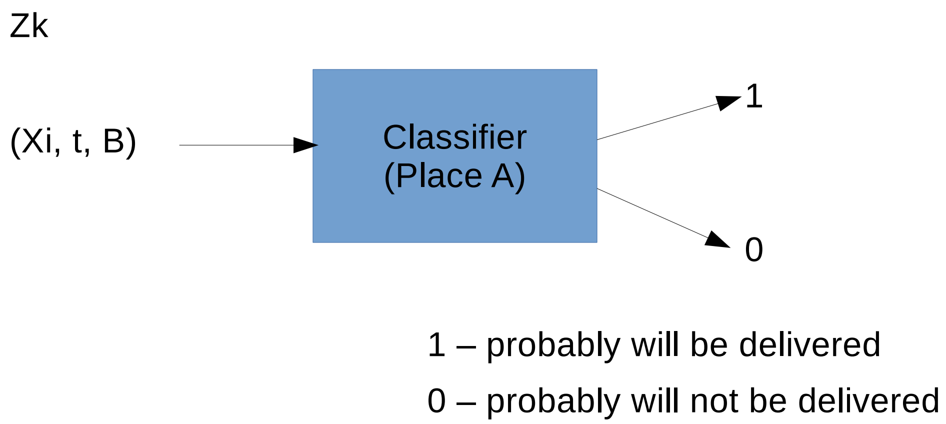

Figure 3.8: A binary classifier for the prediction of message delivery

binary classes: 1 = MESSAGE DELIVERED (forward a message), 2 = MESSAGE NOT DELIVERED (keep a message) and the message delivery delay ¡ μ, μ is the predefined value of m. If the classification for different values of m is required, we can construct a set of classifiers, each of them trained for the different value of maximum message delivery delay. Another approach is to construct a classifier trained on labeled quadruples (X_i, t, B, m).

In a next step, we infer the proposed routing schema from the stationary places A,B to mobile nodes. Let X_k be the OPN node, which is in communication distance of of X_i at time t and simultaneously, place A is in communication distance of of X_i at time t. Assume X_k generates a message MSG addressed to B instead of the place A. From the viewpoint of routing between X_i and B at t, there is no difference between A-X_i... B trace and X_k—X_i—B. This can be extend to all OPN nodes which are in communication distance of X_i at time t. Similarly, we can extend the destination from B to the set DestPlaceNodes(B) of all nodes, which are located in a communication distance of X_j at time t_msg_delivered, where Y_j is the "last hop" node in a communication trace between places A and B.

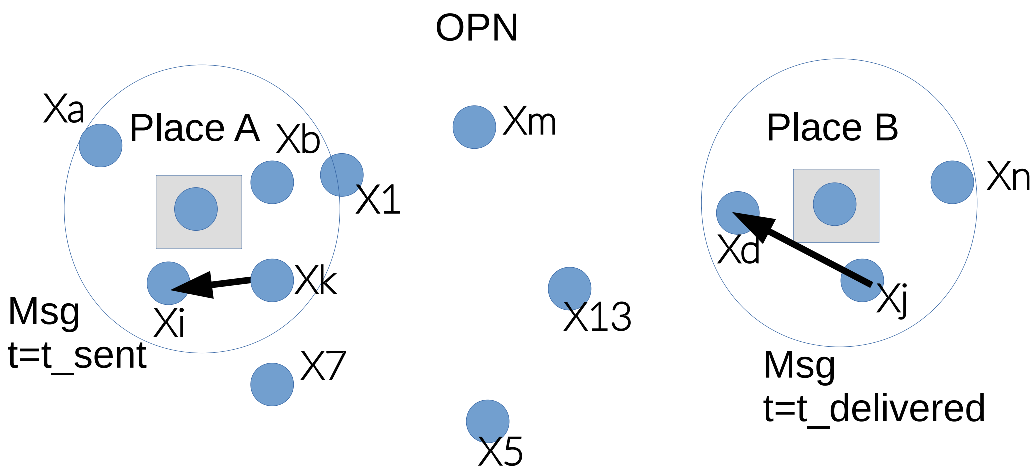

Figure 3.9: Extension of the message delivery prediction method

Training data are extended by the sets DestPlaceNodes(B) at time t_msg_delivered. These sets can be constructed from the node positions data. Let us call the small geographic location around places A, B "cell A", respectively cell B. We discuss below how to compute the size of these cells. The classification task can be reformulated as follows: classify a triple (X_i, t, X_{DEST}), where X_i is the node, which receives a message at the cell A from the source OPN node Z_i, t is the time where the forwarding request occurs, and X_{DEST} is the destination node, to which the message is addressed, into two classes: 1 = MESSAGE DELIVERED (forward a message), 0 = MESSAGE NOT DELIVERED (keep a message) and the message delivery delay ¡ μ, μ is the predefined value of m. It is obvious that the classification is valid only for the source nodes, which move through cell A, and the destination nodes, which move through cell B, not for the whole geographic area of OPN.

It is inconvenient to train classifier at all samples of training data, rather it is better to use short time slots τ. The OPN network is unstationary system. In each time slot, the unstationary system of OPN is approximated by the model of stationary system. The length δ_τ of time slot τ is the parameter of classification and must be selected with the respect to node movement speed, data transmission speed and maximum communication distance. Fig. 3.10 shows the classification scheme after the time slots have been applied to data.

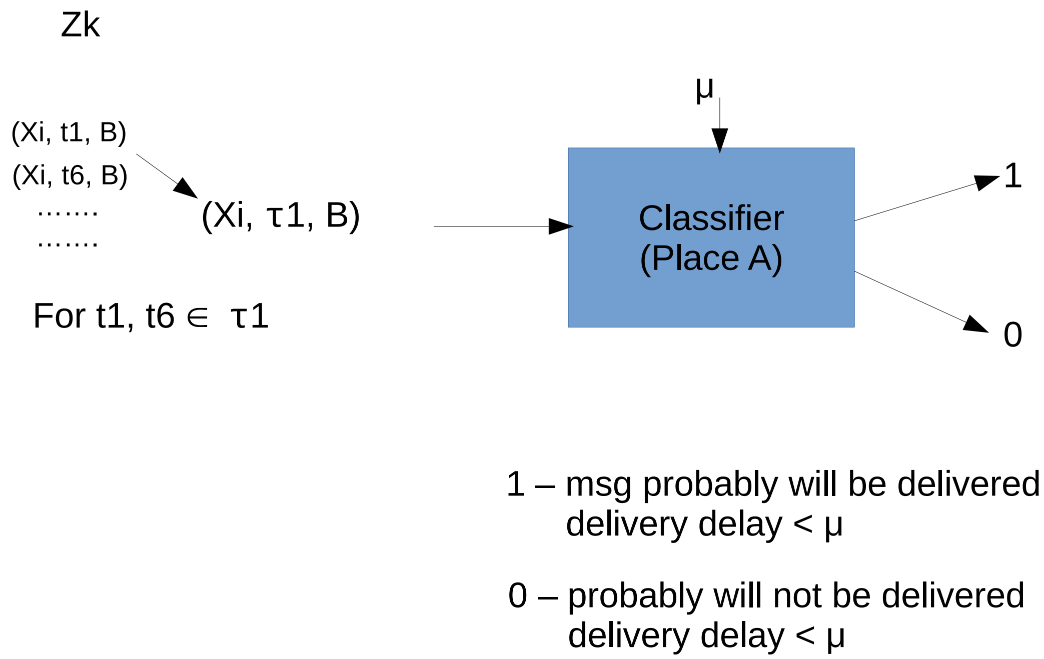

Figure 3.10: The classification scheme after the time slots have been applied to data

Now, we generalize the proposed routing method in order to be this one applicable to all nodes of the OPN. We start by dividing of the OPN geographical area Λ into a grid of cells. The length of a cell diagonal must be equal or shorter than the maximal communication distance of a pair of nodes. This cell size limitation implies that each two nodes inside the cell are in communication distance and can exchange messages.

Consider nodes X_k and X_i which encounter in cell A_{xy}. X_k carries or generates message MSG addressed to node $X_{DEST}, X_{DEST} \in \Lambda$ (e.g. X_{DEST} can be located anywhere in the geographical area Λ). The classification task is extended as follows: classify a triple (X_i, t, X_{DEST}), where X_i is the node, which receives a message at the cell A from the source OPN node Z_i, t is the time where the forwarding request occurs, and X_{DEST} is the destination node, to which the message is addressed, into two classes: 1 = MESSAGE DELIVERED (forward a message), 2 = MESSAGE NOT DELIVERED (keep a message) and the message delivery delay ¡ μ, μ is the predefined value of m. The classification scheme is shown on Fig. 3.12. The task can be solved by the array of N-1 classifiers affiliate to cell A_{xy}, where N is the number of cells in the geographical area Λ. The total number of classifiers is

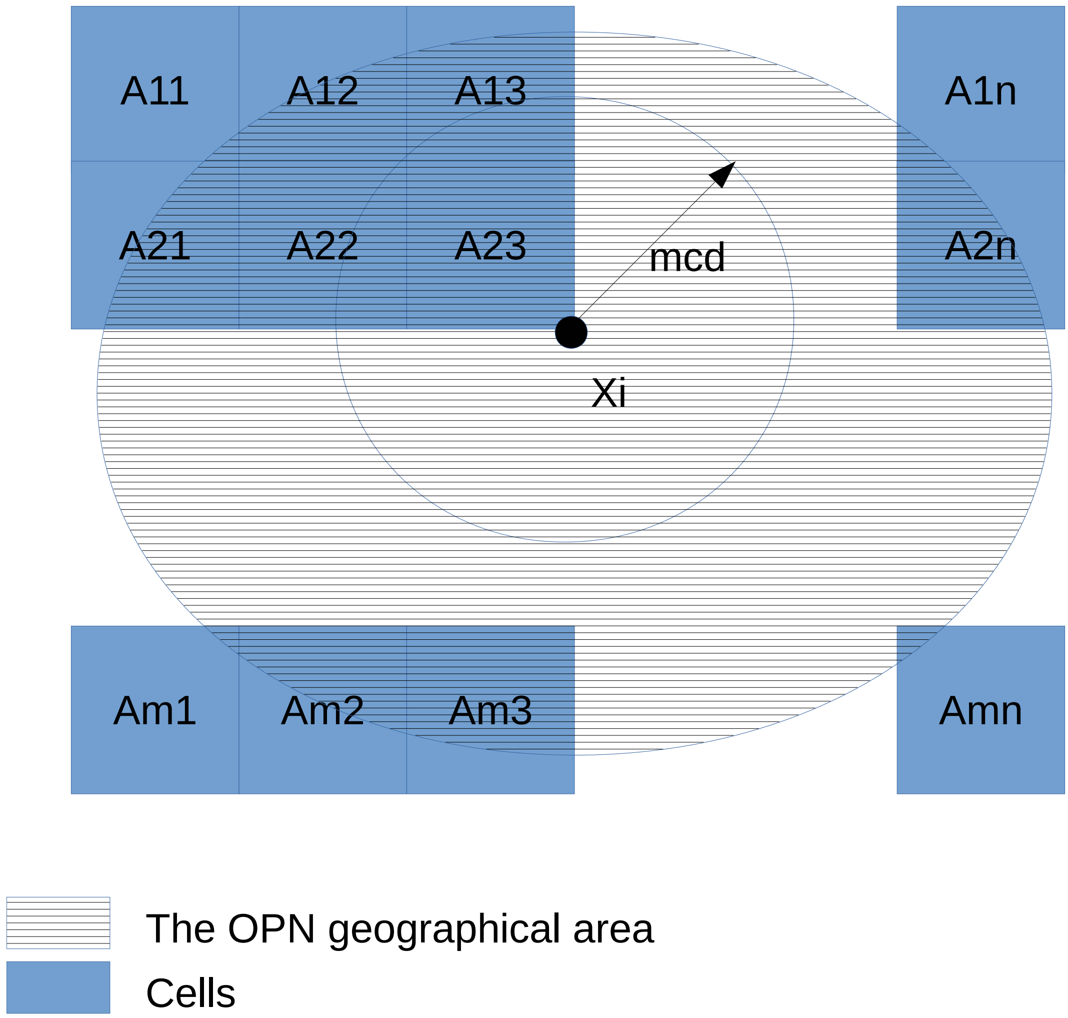

Figure 3.11: Dividing the geographic area of OPN into cells

N(N-1). For each cell, the set of classifiers can be replaced by one classifier, but it can be more convenient to train a set of simple classifiers than the complex one.

3.6.2 Supervised Machine Learning Method Selection

In the literature, many classifiers using different supervised machine learning methods have been described. We recapitulate the requested properties of the classifier n order to work properly in our classification task: i) binary classification, ii) scalability and iii) ability to classify large input data sets. Support Vector Machines (SVM) fulfill all these requirements. Support Vector Machines are supervised learning algorithms introduced by Vapnik [207], [208]. Given a set of labeled training examples, they can be trained to perform non-probabilistic binary linear classification. A Support Vector Machine constructs a hyperplane or set of hyperplanes in a high dimensional or infinite dimensional space in order

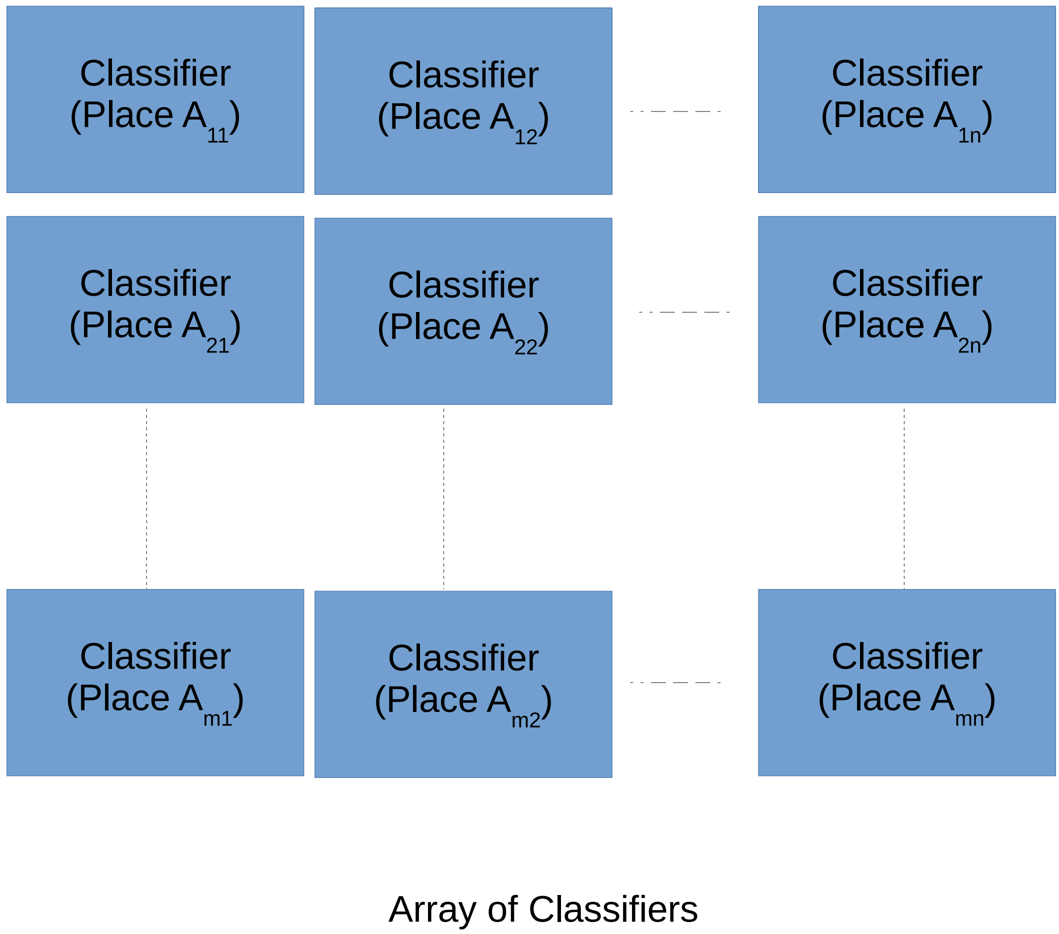

Figure 3.12: A set of classifiers for the geographic area of OPN divided into cells

to separate the labeled data. The detailed mathematical description of the Support Vector Machines is out of scope of this thesis and can be found in [207], [208], [198], [45], [72].

3.6.3 SVM-based Routing Metrics Inference

This section describes step-by-step the process of SVM-based Routing Metrics inference. The process is schematically shown on Fig. 3.13 .

3.6.3.1 Simulation Setup and Dataset

Dataset obtained from node movement simulation. After the simulation scenario is prepared, we run the simulation in the simulation environment ONE [91]. The output of simulation is a file containing records of node positions at each time step. The records have the following form:

$t_i, x_1, y_1, x_2, y_2, x_3, y_3 \ldots\ldots\ldots x_N, y_N$

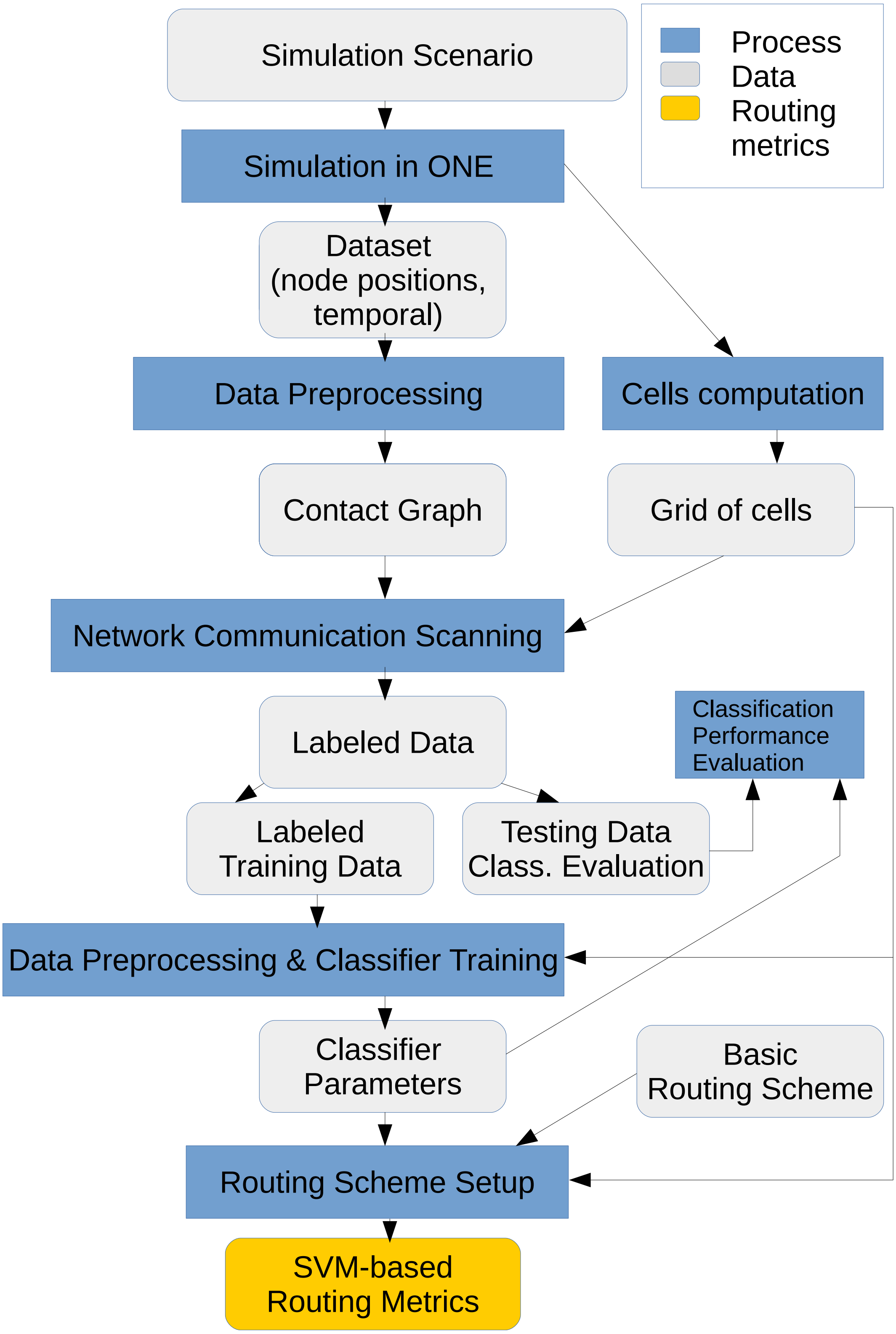

Figure 3.13: SVM-based Routing Metrics Inference: The Block Diagram

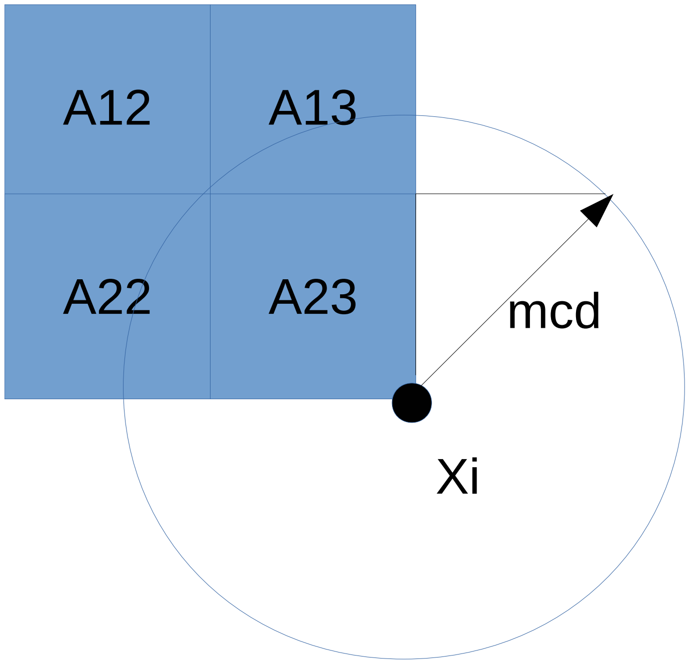

Figure 3.14: Computation of the side of the square cell A

where t_i is a time step, x_j, y_j are the Cartesian coordinates of the node X_{ij} and N is the number of nodes. This records represent a dataset containing temporal data on node positions.

3.6.3.2 Contact Graph

We compute the contact graph of nodes at each time step. The vertices represents the nodes, the edges represent contacts. If two nodes A and B are in communication distance at time t, the contact graph for time t contains an edge between A and B.

3.6.3.3 Grid of cells

Let M be a matrix mxn, which represents the geographic coordinates in the OPN geographical area. We divide this area into N cells. Each cell is a square. The length of cell side is:

$$l_{cell} = \sqrt{max\_com\_distance^2}/2 \tag{3.7}$$

Each place x_i, y_i belongs just to one cell. Each place x_i, y_i has assigned a label of the cell. The matrix mxn can be divided into k x l square sub-matrices.

3.6.3.4 Dataset obtained from communication simulation

In this step, we simulate message forwarding in OPN. We write a simple simulator. The inputs are: temporal contact graph of the simulated OPN, message routing scheme, message injection scheme, grid describing the partitioning of a geographical area of OPN into cells. The messages are transmitted when the nodes encounter. We simulate message forwarding between each pair of cells. Consider the simulation of communication between the nodes in geographic cells A_{mn} and $A_k l$. At time t, the messages are injected into system by nodes, which enter the cell A_{mn} at time t. The messages are injected periodically, the message injection period is a parameter of simulation.

The output of simulation is a file containing records of messages transmitted via simulated OPN. Each record has the following form:

(t_msg_sent, A_{mn}, A_{kl}, X_i, DEST = $\{X_j\}$, t_msg_delivered)

where:

t_msg_sent … time, when the message was sent
A_{mn} … source cell
A_{kl} … destination cell
X_i … source node in A_{ij}
DEST = $\{X_j\}$ a set of destination nodes in A_{kl}
TMD = {t_msg_delivered(X_j)} the set containing time when the message was delivered for the destination node X_j in A_{kl}

3.6.3.5 Training Data Preprocessing and Training Classifier

Labeled data are divided into training data and testing data. Labeled training data are pre-processed, they are filtered by time window. Maximum acceptable message delay is set to μ. The value of μ is selected a classifier training parameter. Classifier accepts (A_{mn}, X_i, τ_k, X_j, CLASS LABEL). A_{mn} denotes cell, where nodes encounter. CLASS LABEL is a binary: 1 = message delivered in time, 0 = message not delivered in time. After the classifier is trained on training data, labeled testing data is used to evaluate classifier performance.

The last step of classifier training phase is the performance evaluation of classification. The evaluation is performed on testing data. The testing data are obtained for the same simulation scenario, but may differ from training data. The classifier performance evaluation for particular scenarios is described in Chapter 4.

3.6.3.6 Routing Metrics Inference

In this step, the routing metrics is constructed. The SVM-based classification routing metrics

- If SVM_Classification (A_{mn}, X_i, τ_k, X_j) = 1, then FORWARD msg (or COPY msg)*)
- If SVM_Classification (A_{mn}, X_i, τ_k, X_j) = 0, then KEEP msg.

*) The routing algorithm combines basic routing scheme and SVM based classifier. Application of FORWARD or COPY depends on the selected basic routing scheme combined with classifier. If the selected basic routing scheme is "forwarding routing scheme", the message is forwarded. If the selected basic routing scheme is "flooding routing scheme", message is copied.

3.6.4 Routing Method

When the classifier is trained centrally, the nodes can upload classifier parameters into routing schemes implemented in the nodes and start forwarding messages using SVM-based routing scheme. This section describes the routing scheme in detail.

MAKE DECISION PROCEDURE PSEUDOCODE

makeDecision (simtime, A, B, DEST)

1. if SVM_Classification(A_{ij}, B, timeSlot(simtime), DEST) == 1 **then**
  return COPY & TIMEOUT;
else
  return KEEP;
end if

This procedure is called by the routing algorithm described in subsection "HRC2 Routing Method". The message exchange policy is same, only the routing metrics differ.

3.7 Routing based on GMRF and Active Node Behavior

3.7.1 Overview of the Method

Problem: to propose an enhancement of routing algorithm using statistical node mobility models and the active node behavior and examine the impact of this approach on communication in opportunistic networks. The active node behavior means that the node itself actively changes its route in order to get to the location more suitable for forwarding the message.

3.7.1.1 Adjacency Matrices and GMRF Models of Node Encounters

Let O be the OPN formed by the set of nodes $X = X_1, X_2 \dots$ in geographical area Λ. The nodes are humans or vehicles. Let $c_{X_i,X_j}(t)$ denotes binary encounter function, which returns 1 if nodes X_i and X_j are in communication distance at t, else returns 0. The matrix $A(t)$ of c_{X_i,X_j}, NxN, where N is a number of nodes, represents one state of a temporary graph $G(t)$ and it is called adjacency matrix. The values of elements of this matrix are zero or one. Let τ_k denotes a time window of length δ and let $C_{X_i,X_j}(\tau(i))$ denotes a counter of encounters of nodes X_i and X_j during the time window τ_i. Similarly, we can define adjacency matrix $A_2(\tau_k)$ for time window tau. The set of adjacency matrices $\{A_2(\tau_k)\}$ can be represented by the grey-scale images of NxN points. Similarly, the set of adjacency matrices $A(t)$ can be represented by the set of black&white images. We can use GMRF in order to construct a simple node mobility model from these adjacency matrices. If the behaviour of OPN changes significantly, the GMRF models change too.

3.7.1.2 Statistical Message Delivery Model

Let the the division of the OPN geographical area into p x q square cells as it has been proposed in section 4.6 SVM-based routing. Consider the network scanning as it described in 4.6 SVM-based routing.

The output of message transmission simulation is a file containing records of messages transmitted via simulated OPN. Each row represents one message. We can construct *the cell delivery probability matrix* matrix $\Pi(\tau)$, N x N for a time window τ where N is number of cells, and the matrix elements are numbers of successfully delivered messages from cell A_{ij} to $cellA_{kl}$ during τ. Fig. 3.15 shows OPN in two states, which differ in temporal node mobility pattern in use. Although the node density and locations are similar, the cell delivery probability matrices differ.

We can use techniques used in computer vision to model cell delivery probability matrices, particularly GMRF model for texture classification. If the behaviour of OPN changes significantly, the GMRF model changes too.

3.7.2 Node Mobility Model based on GMRF

We suppose that the node "snaps" constructed in a way described above can be described by the finite set of 2D GMRF. This simplification needs an assumption about Gaussian texture character of the data patterns. In fact, if there some macroscopic structures were present in input data patterns, it would not be sufficiently modeled by the GMRFs. GMRF is well know technique for the 2D data modeling. Its application for image processing was described by K. Deguchi [48] in or Chellapa [35]. The basic algorithms for estimating GMRF parameters are: *Least Squares (LS) Estimate of GMRF Parameters, Maximum Likelihood Estimate (MLE) of GMRF Parameters.* We follow the approach published by Chellapa [35];

Let $y(m; n)$ be the intensity of an image at pixel (m; n). We suppose that the snap representation of dimension M x M is

$$y(s), s \in \Psi, \Psi = \{s(i,j) : 0 \leq i,j \leq M-1\} \tag{3.8}$$

We assume that the zero mean observations from the given snap y(s) are Gaussian. The GMRF is stationary non-causal two dimensional autoregressive process described by the following equation:

$$\begin{aligned} y(s) &= \sum A_r(y(s+r) + y(s-r)) + e(s) \\ E\big[e(s)e(r)\big] &= -A(s-r)\beta I_{N^*}(s-r) \end{aligned} \tag{3.9}$$

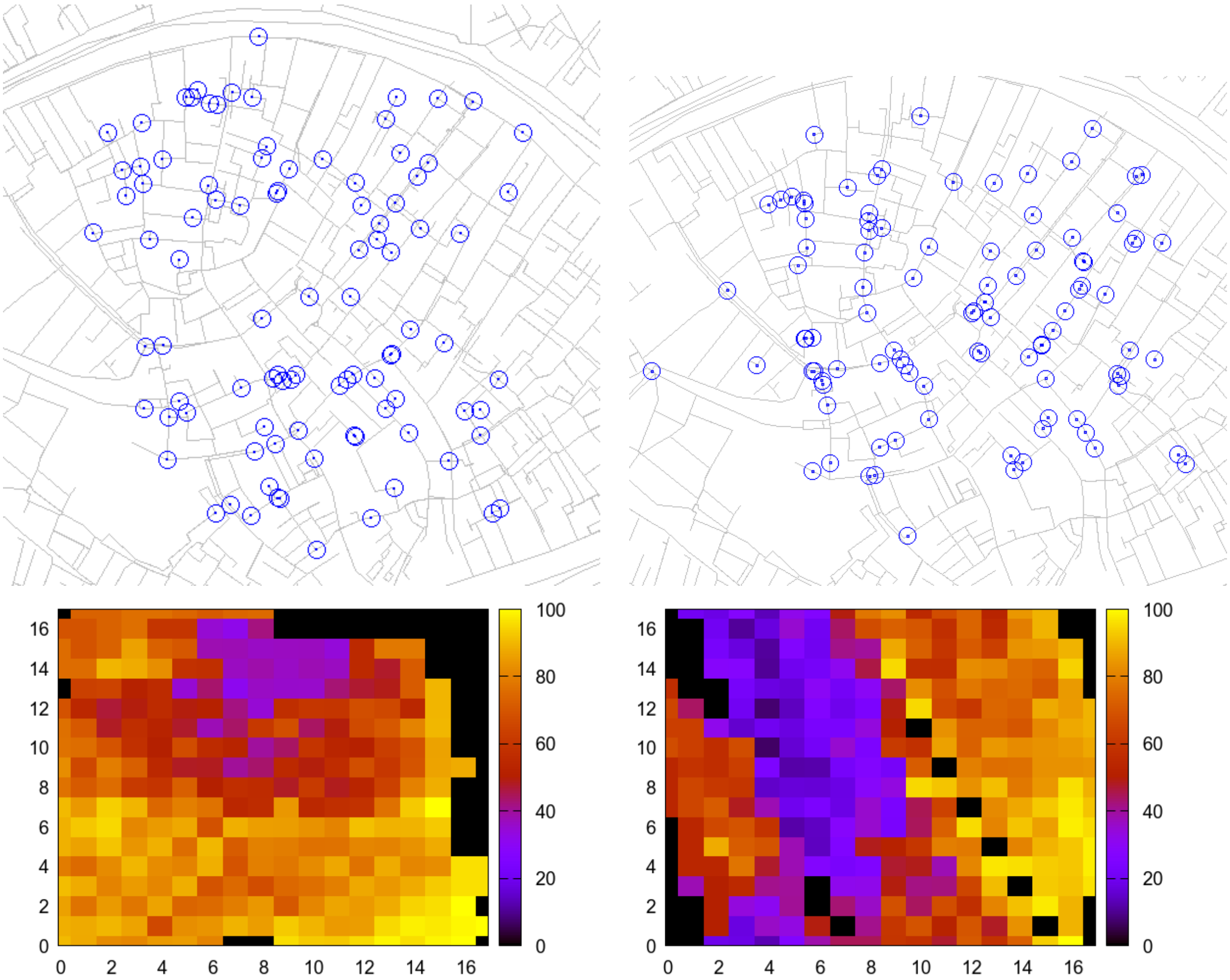

Figure 3.15: An opportunistic network in two states corresponding to different temporal mobility patterns: node positions in a geographical area and computed delivery probability matrices in a graphic form

The zero mean sequence $e(s)$ is correlated innovation sequence. N^* is equal to 1 when $r \in N^*, N^* \in N \cup (0,0)$. N is a set of the neighborhood pixels. The unknown parameters $A = (A_r, r \in N)$ and β can be estimated using LS algorithm or MLE algorithm [172].

Least Squares (LS) Estimate of GMRF Parameters

The snap is described by the equation 3.4 and the GMRF is described by 3.5. The unknown parameters $A = (Ar, r \in N)$ and β can be estimated using LS algorithm as follows [35]:

$$A^* = \big[\sum_{\Psi_i} y_s y_s^T\big]^{-1} \big[\sum_{\Psi_i} y_s y(s)\big] \tag{3.10}$$
$$\beta^* = \frac{1}{M^2} \sum_{\Psi_i} \big[y(s) - {A^*}^T y_s\big]^2$$
$$y_s = y(s+r) + y(s-r), r \in \mathbb{N}$$

Ψ_i is defined as $\Psi_i = \Psi - \Psi_B$, where Ψ_B is a boundary set:

$$\Psi_B = \{s = (i,j), s \in \Psi, (s+r) \notin \Psi\} \tag{3.11}$$

for at least one r. The LS estimates of the parameters A^* and β^* make a feature vector, which characterizes the analyzed snap.

Maximum Likelihood Estimate (MLE) of GMRF Parameters

Correlations over the window defining a model are sufficient statistics. We compute statistically sufficient representation of the observed texture. Similarly to LS Estimation, MLE estimation is standardized way of GMRF parameters estimation.

$$E\big[e(s)e(r)\big] = -A(s-r)\beta I_{N^*}(s-r) \tag{3.12}$$

Then we can compute the conditional probability function $p(y|A,\beta)$.

$$p(y|A,\beta) = \frac{|H(A)|^{1/2}}{(2\pi\beta)^{M^2/2}} \exp\left(\frac{-y^T H(A) y}{2\beta}\right) \tag{3.13}$$

where $H(A)$ represents the transformation matrix $H(A)y = e$. The sample correlation function is defined as follows:

$$C_d(r) = \frac{1}{M^2} \sum_{s \in \Psi_i} y(s) y(s+r) \tag{3.14}$$

The quadratic form

$$y^T H(A) y \tag{3.15}$$

can be simplified as:

$$y^T H(A) y = M^2 \big[C_d(0) - A^T C_d\big] \tag{3.16}$$

where $C_d = \text{col.}\big[C_d(r), r \in \mathbb{N}\big]$. Then we can rewrite the equation for conditional probability density function:

$$p(y|A,\beta) = \frac{|H(A)|^{1/2}}{(2\pi\beta)^{M^2/2}} \exp\left(-\frac{M^2}{2\beta}\left\{C_d(0) - A^T C_d\right\}\right) \tag{3.17}$$

Then we can use the Nyman-Fisher factorization theorem and we receive:

$$\alpha = \{C_d(0), C_d(r) | r \in \mathbb{N}\} \tag{3.18}$$

Then a is a sufficient statistic for (A, β). $\mathbb{N}$ is defined neighborhood.

3.7.3 Routing Metric

3.7.4 Active Node Movement Algorithm

This section deals with the design of the Active Node Movement Algorithm using the statistical node mobility model and its implementation into node routing algorithm. The active node movement means that the node itself actively decides to change its route to increase the probability of message delivery.

The implementation of motion of nodes is the part of the simulation environment. The node route may be fixed (for example, bus lines). The route of the node doesn't need to be precisely defined, only the target coordinates can be defined. The nodes then move on the shortest routes. The simulator calculates the routes using the Dijkstra algorithm.

The node has a limited message buffer and can not transmit any number of messages. There are also limitation set to the node deviation from its planned route. In the real world, for example, a driver or a pedestrian would change the route only once or twice during a certain period of time. In the real world, the pedestrian or vehicle driver would only deviate from the planned route within a certain distance.

We designed the Active Node Movement Algorithm (ANMA) for implementation in two parallel threads. In the first thread, decisions are made about message forwarding and active deviation from the route. In the second thread, the algorithm for not exceeding the maximum number of deviations of the node from the planned route within a given time interval is implemented.

ANMA PSEUDOCODE

gmrfNetState()

1. if $GMRF\_NET\_STATE == DIVIDED$ **then**
  if MOVEMENT == yes **then**
    oldTarget = getTarget();
    MoveTo(GenerateNewTarget());
    MoveTo(oldTarget);
  else
    // continue to old target
  end if
else
  // continue to old target
end if

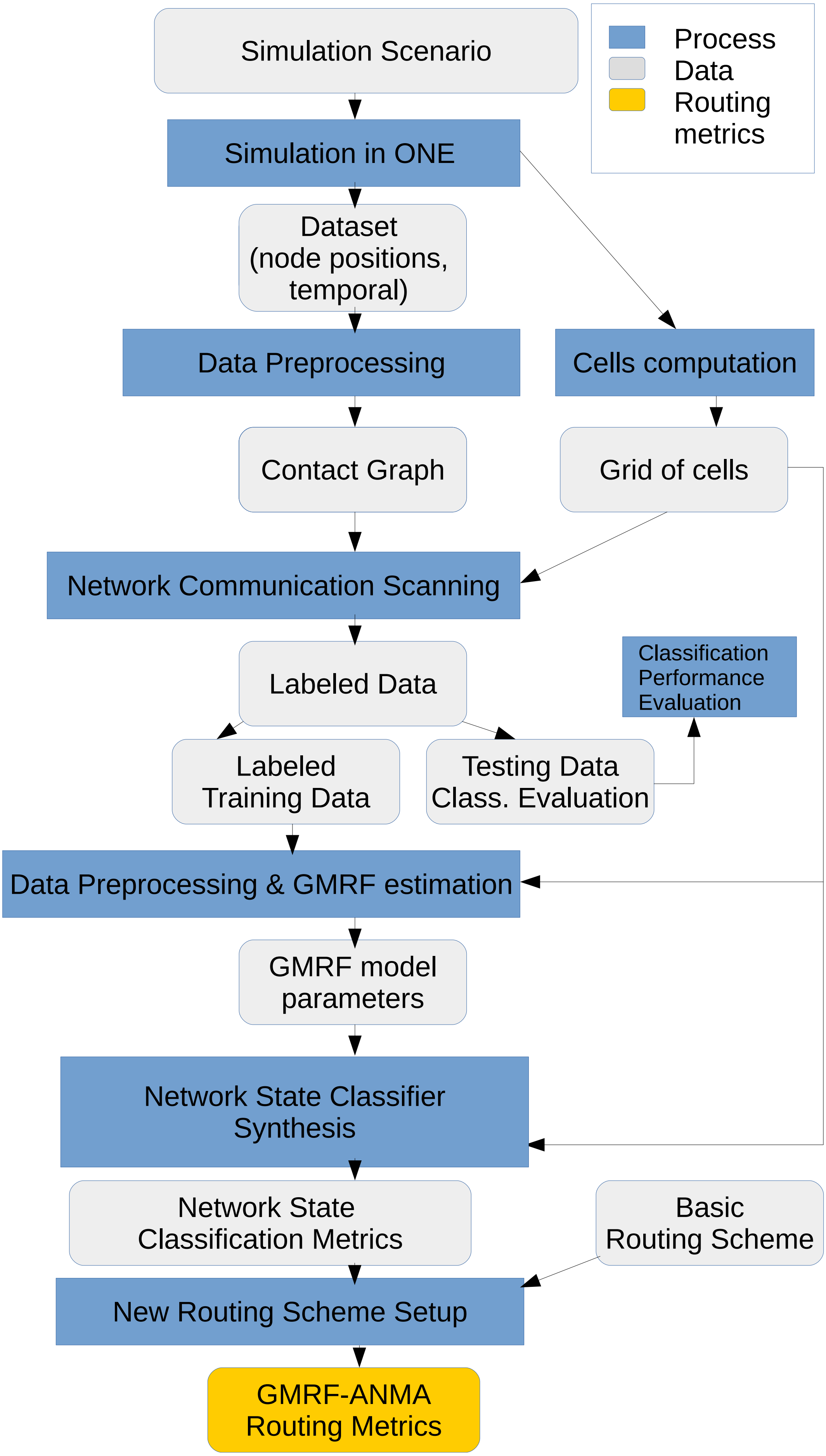

Figure 3.16: GMRF and ANMA routing metrics inference $A(\tau_k)$

Chapter **4**

Main Results

4.1 ONE: Opportunistic Network Simulation Environment

The simulations and performance evaluation were analyzed through node mobility simulation using the Opportunistic Network Environment (ONE) simulator [93], which has been reported previously as a simulation environment in scientific literature on OPN routing protocols. Using the ONE simulator, Li et al. [111] have studied how the selfish behaviors of nodes affect the performance of DTN multicast. They used standard mobility model available in the ONE simulator. Socievole et al. compared six different routing protocols using simulation scenario with random way-point mobility model in the simulator ONE[185]. Spaho et al. [189] conducted simulations with the ONE simulator in order to evaluate and compare the performance of four different routing protocols in a many-to-one communication opportunistic network. In [64] the simulator ONE has been used to evaluate the performance of SRAMSW routing algorithm.

ONE is an agent-based discrete event simulation engine. The main functions of ONE consists of i) modeling of node movement, ii) modeling of inter-node contacts, (3) modeling of message handling and (4) modeling of routing. In addition to these main functions, ONE offers data post processing tools, visualizations tool and report generating tool. A detailed description of the simulator is available in [91] and the ONE simulator project page [92] where the source code is also available.

Fig. 4.1 shows the structure of the ONE simulation environment.

ONE is designed as the open source software for evaluation of the routing algorithms. It has implemented function of adding of a new routing algorithm (a creation of a new routing module). This feature enables to implement and test user's own routing algorithms in ONE. ONE has been implemented in Java and it is available under the GPL. ONE is designed to be able to interact with other programs and data sources. The simulator has implemented interfaces for some main functions as node movement, connectivity or message routing traces. ONE supports interaction with external data processing software via report module. A report module can communicate in real-time with external software [93]. This feature enables use user's own code or some external software tools.

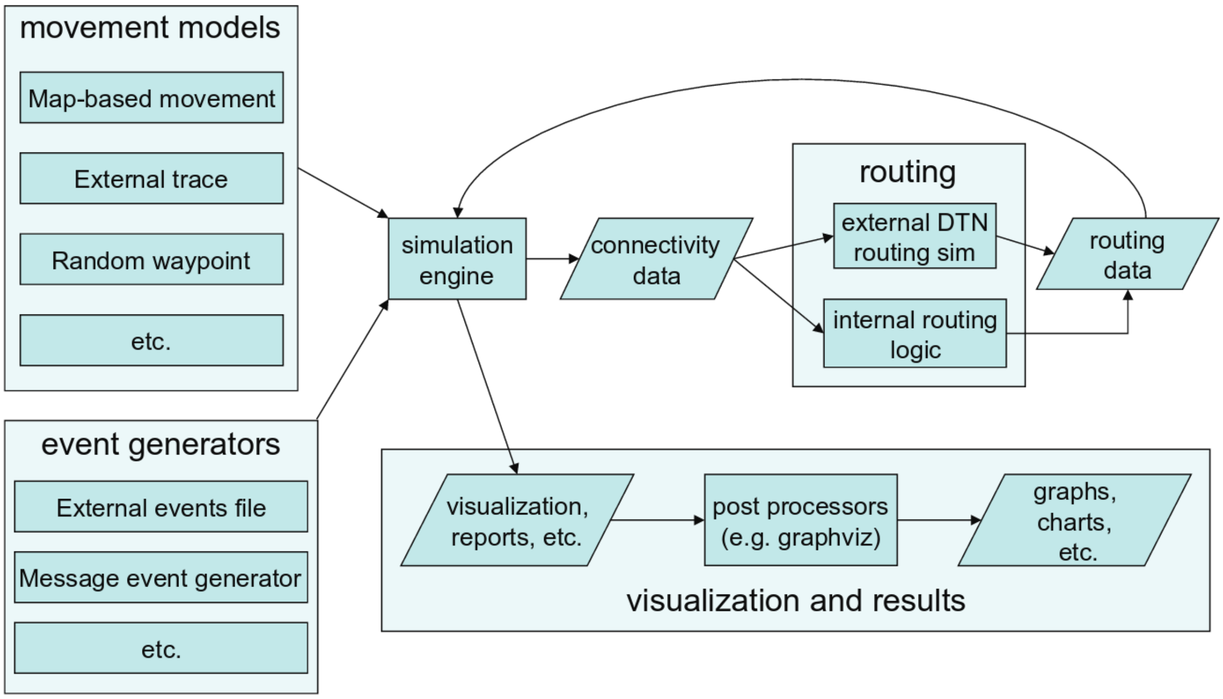

Figure 4.1: The structure of the ONE simulation environment [93].

4.2 Performance Metrics

We adopted the following metrics to evaluate the performance of the proposed algorithms and to compare them to other selected routing methods.

Message Delivery Ratio. It is computed as the ratio of the delivered messages to the number of all of transmitted messages multiplied by 100. The larger values of message delivery ratio imply the better routing protocol performance.

Overhead Cost Ratio. It is computed as the number of transmitted messages in the network divided by the number of all created unique messages. This performance metrics reflects the efficiency of the evaluated routing protocol.

$$OCR = \frac{number of transmissions}{number of created unique messages} \tag{4.1}$$

Average Message Delivery Delay. The lower values of average message delivery delay imply better routing protocol performance.

Average Number of Hops of Delivered Messages. Average Number of Hops is the total number of hops of successfully delivered messages divided by the number of delivered messages. It characterizes the process of message forwarding through OPN.

4.3 Routing Methods for Comparison

We selected four protocols were tested in the simulation: i) First Contact (FC), ii) Epidemic (EP), iii) PRoPHET, iv) BUBBLE-Rap.

First Contact routing protocol is a forward-based routing strategy; only one copy of a message exists at a certain point in time in a system. The node that carries a message, forwards this message to the first node, which it encounters, unless the message is transmitted to the destination node.

Epidemic Routing is flooding-based routing scheme. It is well-known and usually it is considered a reference for other routing methods.

PROPHET is probabilistic routing scheme, which improves routing using routing metric based on probabilities of node contacts computed from the node contacts in history.

BUBBLE-Rap is a community-aware routing scheme, which combines the node centrality routing and node labels based on node affiliation to community. Authors used two methods to detect communities: k-clique community detection algorithm proposed by Palla [146] for the detection of overlapping communities, and a modularity based approach proposed by Newman et.al [142]. We implemented k-clique based version of BUBBLE RAP. We implemented a version with centrally computed communities in order a method could be compared to proposed routing schemes which infere routing metrics centrally.

4.4 Experiments

In order to work with simulation scenarios, which are more close to to real-world human mobility, we proposed the following method for simulation setup generation:

1. The geographical area selection: user defines boundaries of geographical area, in which are located targets (low density residential, high density residential, industrial targets, commercial targets, school & universities, ...). The selected geographical area of OPN may not be necessary continuous, if two constraints are fulfilled: i) all separated areas must be included in the geographical area Λ of the analyzed OPN, ii) there is traffic infrastructure available for node movement among areas. The selected area (areas) are approximated by set of zones (S_i, r_i), where S_i is the center of i-th zone, and r_i is the radius of the i-th zone.

2. User defines the number of categories of targets.

3. User defines probability distribution of placement of target localities in each zone (Gaussian, uniform).

4. Targets are generated in accordance to parameter setup from steps 2 and 3. Targets can be also added manually.

5. User defines a number of nodes for each zone, which have their "home places" in the zone. Then, the home places are assigned to nodes. It can be done automatically or manually if it is needed for simulation purposes.

6. User defines categories of nodes with the similar movement pattern. Movement patterns consists of the set of mobility sub-patterns and time schedule (time intervals, when each mobility sub-pattern is applied). The sub-pattern is defined by the following parameters: i) number of targets, mobility "motif", distance and distribution of targets, time-to-stay in target)*. User defines the affiliation of the nodes of different categories.

7. The probability distributions of time schedule divergences are defined.

8. The targets are imported to the simulator ONE. The initial node positions (= home places affiliated to the nodes are imported in the simulator ONE. The file describing node mobility model using simple commands is imported by the simulator ONE. The proposed time schedule is recomputed to be implicitly present in the commands *). The explicitly defined times*) in commands are recalculated with the respect to the probability distributions defined in step 7: the randomly generated divergences are added to or subtracted from the defined time values*).

*) It corresponds to real-world human mobility. For example, one can leave his home each morning approximately at half past seven, but in fact he leaves his home between 7:25 and 7:35. This function was implemented to make the simulation scenarios more real.

4.4.1 Experiment 1

Simulation Scenario 1: Geographically structured OPN
Our tested method: Hierarchical routing with clustering 1 (hrc 1)
Compared to: Epidemic Routing, PROPHET, First Contact, Bubble Rap

Proposed experiments:
Influence of a period of message generation to Performance Metrics
Influence of TTL to Performance Metrics
Influence of Buffer size to Performance Metrics
Influence of Geographical Structure of Network to Performance Metrics
Influence of Time dependent connectivity between two OPN networks to Performance Metrics
Influence of Changing Node Mobility Models during Simulation

OPN Setup:
Urban area, road density:high
The number of separated Target Geographical Regions: 5
Analyzed Simulation Interval 8 hours from each day of 10 day simulation
Changing Mobility Patterns: 1) day-night model, 2) day changing mobility patterns inside each region
Probabilistic distribution of targets in regions: uniform

Probabilistic distribution of node initial positions in regions: uniform
Probabilistic distribution of node stays in targets during the day phase: log-normal

Parameter	Range
Map	Venice
Simulation size	4500 x 3400 m
Moving speed	random 0.5 - 1.5 m/s
Transmission range	20 m
Simulation Time	432000 time units ~ 10 days
Sampling Period T_s	0.5 time units ~ 1 second
Message size	36 bytes
Node Buffer Size	1 - 500 messages
Message Generation Period	10 - 500 time units ~ 20 - 1000 seconds
Time to live	1 - 300 transmissions

Table 4.1: Simulation Scenario 1: Experiment setup for ONE Simulator

This experiment has been conducted on simulation scenario 1. The OPN consists of 5 separated target regions located in the urban area of Venice (high density o roads), with uniform probabilistic distribution of targets in each region. We selected the uniform probabilistic distribution of node initial positions in each OPN geographical region and log-normal probabilistic distribution of node stays in targets during the day phase. The simulation was conducted for 10 days with day-night pattern and periodically changing day traffic pattern in each region. For the purposes of OPN routing, we selected the data collected in time interval of 8 hours (7 AM to 15 PM) from each day of 10-day simulation. The data from 4 days were used t train the model. The data from six days were used to test the performance of the proposed methods. The simulation was conducted for 1000 nodes. The simulation results are presented in graphs.

4.4.1.1 Influence of Number of Nodes to Performance Metrics

We conducted ten simulations, each of them with the different number of nodes: 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000. The message node buffer was set to 200 (messages), TTL to 500 time units (1000 seconds) and message generation period to 500 time units (1000s). One simulation time unit is 2s. Other parameters of the simulation were not changed.

RESULTS: Fig. 4.3 shows the message delivery ratio (referred as delivered messages [%] in graphs), average message delivery delay, average number of hopes and overhead cost ratio as functions of the number of nodes in simulation. From this figure it can be seen that there is an almost linear dependency between the number of nodes and message delivery ratio observed for all methods except epidemic routing. Poor performance of epidemic routing is caused by the earlier network congestion. This linear dependency is caused by the

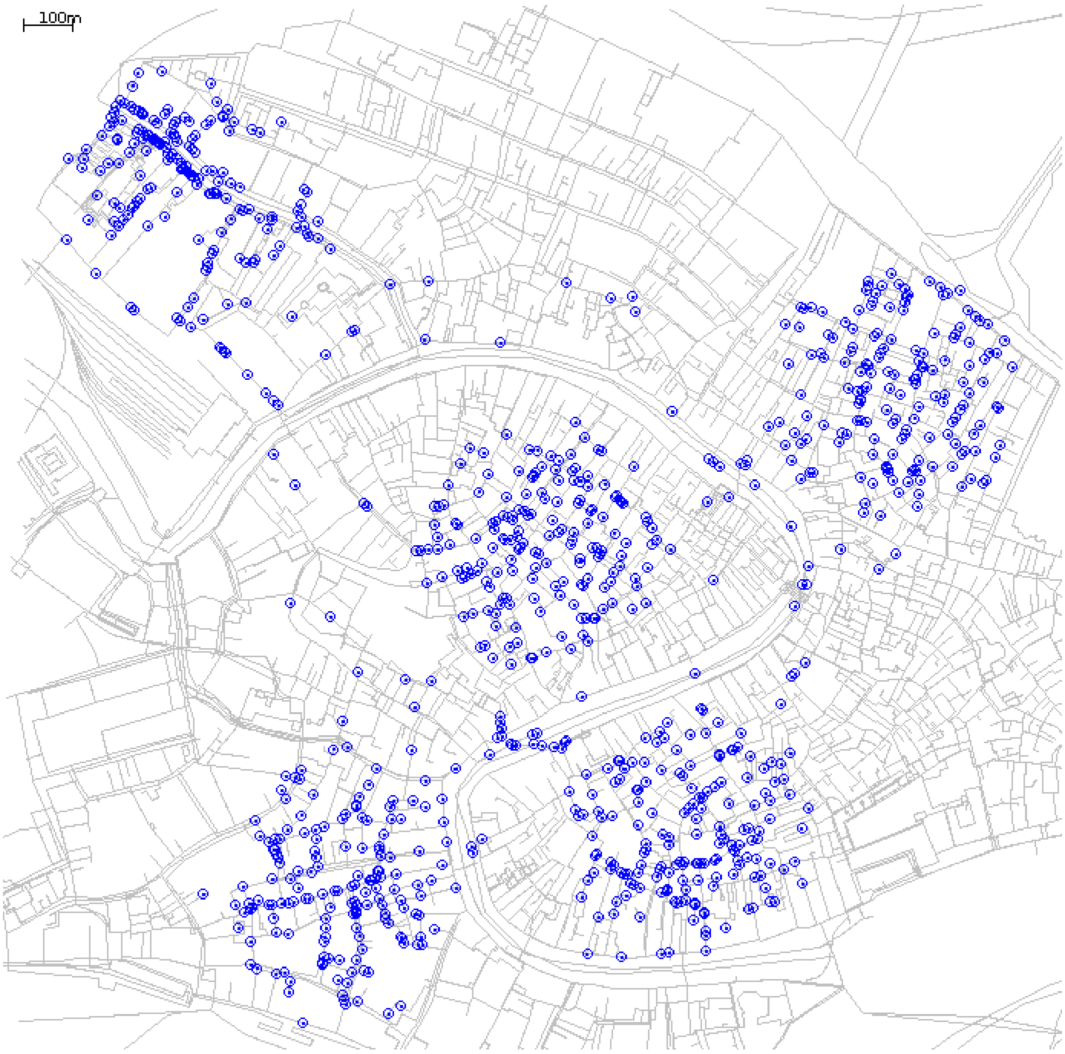

Figure 4.2: Simulation Scenario 1: The initial positions of nodes in OPN geographical area

dependency of successful message delivery on a small set of nodes, which travel to more than one region. As the number of all nodes increases, the number of nodes travelling between regions increases and a higher number of messages is delivered. Furthermore, it can be observed that the proposed routing schema Hierarchical Routing with Clustering (HRC1) outperforms all other methods in the number of delivered messages. It achieves about 50 percent of delivered messages on simulation setup with 1000 nodes, while the success rate of all other methods is in the interval of 15 to 20 percent. The graph of the average message delivery delay as a function of the number of the nodes in simulation indicates that the proposed routing method has the large delay for small number of nodes in simulation except the interval of 100 to 200 nodes. The average message delivery delay decreases as the number of nodes increases. From the viewpoint of average message delivery delay, the BUBBLE-Rap conducts well. The Epidemic routing achieves the shortest average message

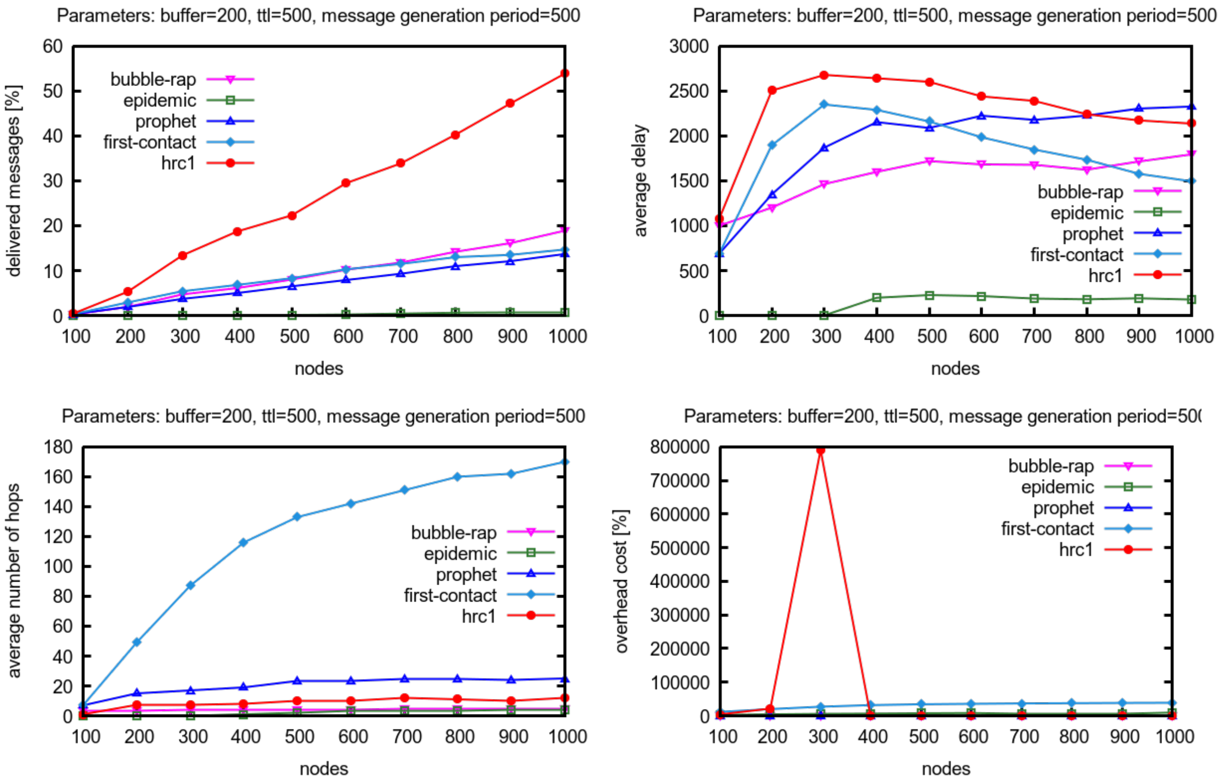

Figure 4.3: Message delivery ratio, average message delays, average number of hops and overhead cost ratio as a function of the number of the nodes in simulation

delivery delay, which is partially influenced by the character of this routing scheme.The average number of hopes rapidly increases when the First Contact routing scheme. It is caused by the random character of its routing scheme. The graph of average number of hopes as the function of the number of nodes is almost flat for the other methods. HRC1 is comparable to BUBBLE-Rap. The last graph indicates that there is an unexpected extremely high peak of overhead cost ratio for HRC1 routing when the simulation was conducted for the number of nodes equal of 300. The configuration with 300 nodes leads to large communities. In HRC1 routing, the routing inside communities is conducted by epidemic routing. The occurrences of peaks of overhead ratio can be eliminated by implementation of more tight constraints of application of epidemic routing into method or by limiting the size of node communities.

4.4.1.2 Influence of Message Generation Period to Performance Metrics

Simulation Setup We conducted several sets of simulations for different values of message generation period and observed the influence of different message generation period to the performance metrics. The range of message generation periods was of 1 to 500. Lower values of message generation period imply higher rates of message generation by nodes, and consequently higher number of messages, which are simultaneously present in simulation. The other simulation parameters TTL and node message buffer were set as follows:

1. set of simulations: TTL = 500, buffer = 500
2. set of simulations: TTL = 500, buffer = 50
3. set of simulations: TTL = 500, buffer = 50
4. set of simulations: TTL = 50, buffer = 500
5. set of simulations: TTL = 50, buffer = 100
6. set of simulations: TTL = 50, buffer = 50
7. set of simulations: TTL = 5, buffer = 500
8. set of simulations: TTL = 5, buffer = 50

Both the TTL and message generation periods are referred in time units of simulation. 1 time unit = 2 s. The influences of message generation period to all four performance metrics were analyzed.

RESULTS:

Fig. 4.4 shows the message delivery ratio (referred as delivered messages [%] in graphs) as a function of the period of message generation. Lower values of message generation period imply higher rates of message generation by the nodes, and consequently the higher number of messages, which are simultaneously present in simulation. As it is shown in graphs, the extremely low values of message generation period cause network congestion. Furthermore, it can be observed that the achieved results strongly depend on the size of message buffer. For the extremely short buffers (buffer = 5), the performance of the proposed method in decreases and it is smaller than 10 percent. The best results were achieved for the large buffer (buffer = 500). The proposed routing schema Hierarchical Routing with Clustering (HRC1) outperforms all other methods in the number of delivered messages for all tested combinations of parameters TTL and buffer. The best results have been achieved for the combination of TTL = 500 time units (1000 s) and the buffer size = 500, and for the combination of TTL = 50 time units (100 s) and the buffer size = 500, where the message delivery rate ratio achieves almost 90 percent of delivered messages. It seems that the highest influence on message delivery ratio has the combination of all three parameters buffer size, TTL and message generation period. We observe two types of dependencies between the message generation period and message delivery ratio. For large values of the buffer (buffer = 500), the number of delivered messages depends on the message generation ratio until the period reaches the 200 boundary; than the network is saturated and the ratio doesn't grow for higher values of message generation period anymore. The second type of dependency can be observed for the smaller values of buffer (buffer=5, buffer=50). The graph indicates a correlation between the message generation period and the observed message delivery ratio.

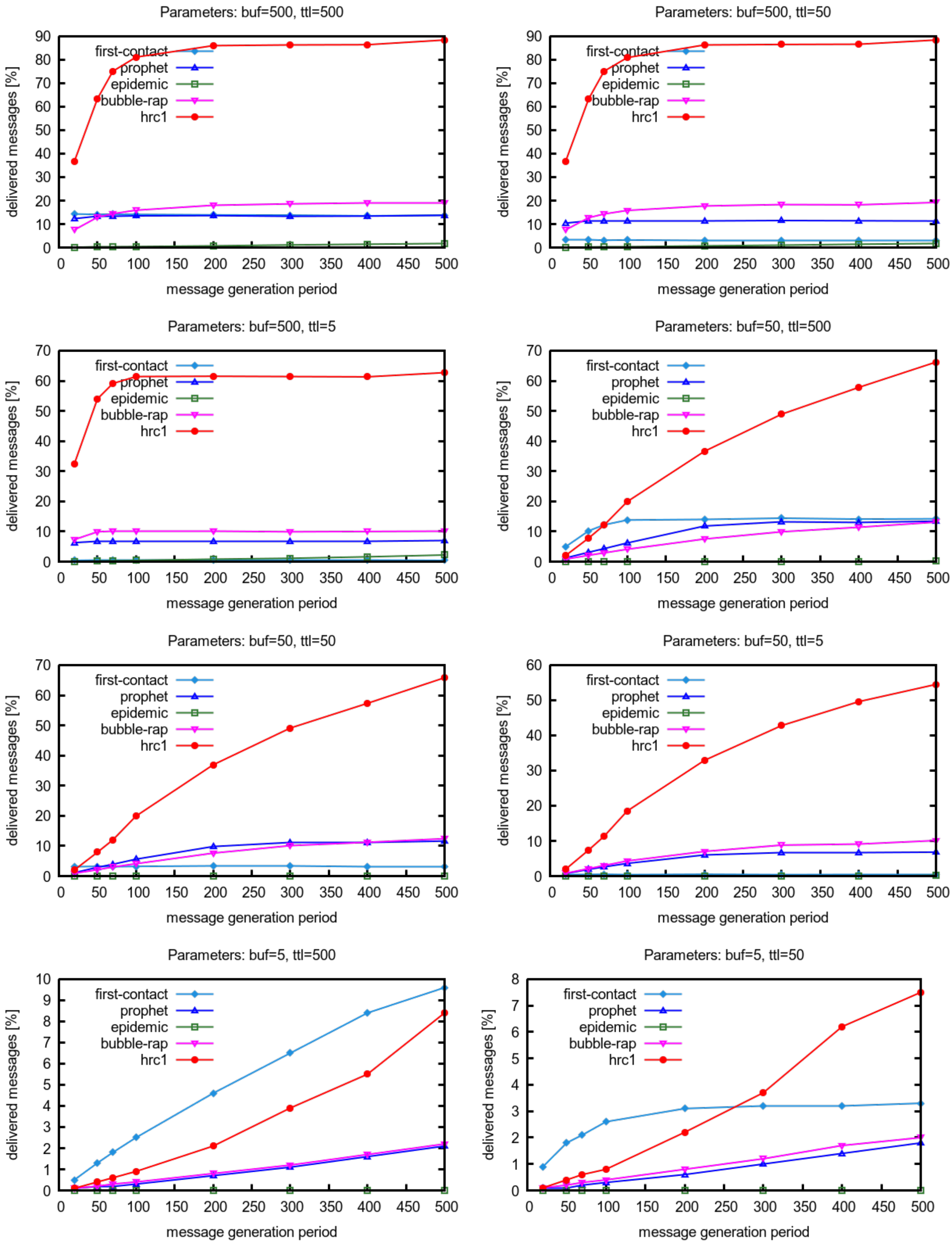

Figure 4.4: Message delivery ratio as a function of the period of message generation

Fig. 4.5 shows the average message delivery delay (referred as average delay in graphs) as a function of the message generation period. Epidemic routing and First Contact routing outperform all other methods. This result is in accordance with our assumptions, because both these methods use simple routing rules and routing is very fast. The average message delivery delay of the proposed routing schema HRC1 is comparable to BUBBLE-Rap for the longer message generation periods. For message generation period = 50 (100 s) BUBBLE-Rap works slightly better. The average message delivery delay of the PROPHET protocol depends rather on TTL than on the size of node message buffer. For extremely low values of TTL, the PROPHET routing scheme achieves the worst results. The proposed routing scheme HRC1 achieves boldly high values of average message delivery delay for the small values of message buffer. The average message delivery delay decreases with the increasing buffer size. The improvement in average message delivery delay, which appears with growing size of buffer, is high for small sizes of the buffer, but as the size of the buffer grows, the dependency on the buffer size becomes weak.

Fig. 4.6 shows the average number of hops as the function of the message generation period. In accordance to our assumptions, the high number of hopes can be observed when the First Contact routing scheme was in use in simulation. The average number of hopes of the proposed routing schema HRC1 is low and it is comparable to BUBBLE-Rap and Epidemic Routing. The graphs are almost flat. It can be interpreted as considering that the number of hopes depends particularly on the applied routing method.

Fig. 4.7 shows the overhead cost ratio (referred as overhead cost in graphs) the function of the message generation period. The overhead cost ratio of the proposed method HRC1 is influenced by the combination of parameters buffer size and TTL. For the buffer=5000, the overhead cost ratio is about 40000 percent. First, second, third and fourth graph refers to high an unexpected extremely high peak of overhead cost ratio for HRC1 routing when. the simulation was conducted for the number of nodes equal of 300. The configuration with 300 nodes leads to large communities. In HRC1 routing, the routing inside communities is conducted by epidemic routing. The occurrences of peaks of overhead ratio can be eliminated by implementation of more tight constraints of application of epidemic routing into method or by limiting the size of node communities.

4.4.1.3 Influence of Buffer Size to Performance Metrics

We conducted simulations for different sizes of message buffer (from 1 to 500 messages) and observed the influence of message buffer size to the performance metrics. The other simulation parameters TTL and message generation period (referred as send period in graphs) were set as follows:

1. set of simulations: TTL = 500, message generation period = 500
2. set of simulations: TTL = 500, message generation period = 100
3. set of simulations: TTL = 500, message generation period = 50

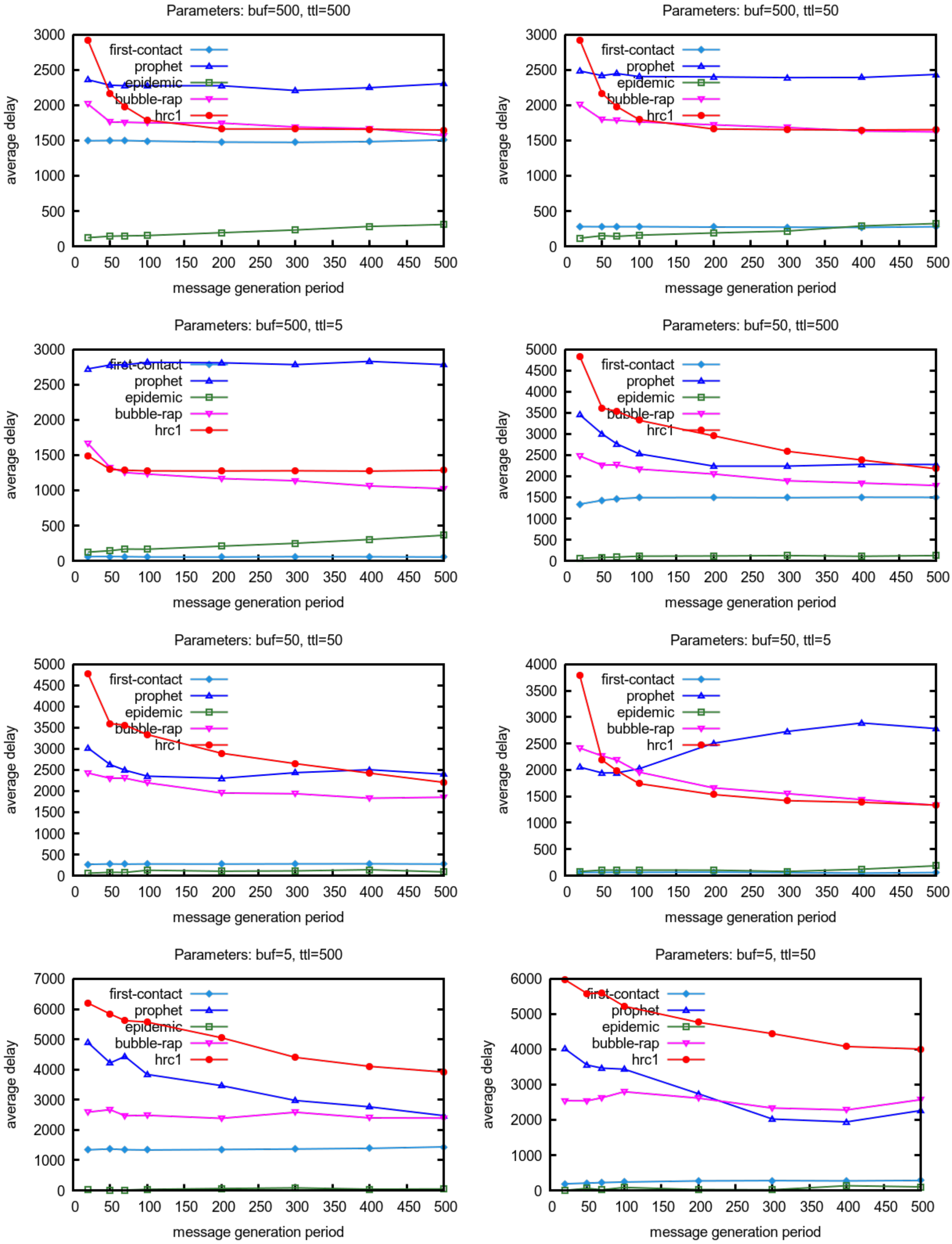

Figure 4.5: Message delivery delay as a function of the period of message generation

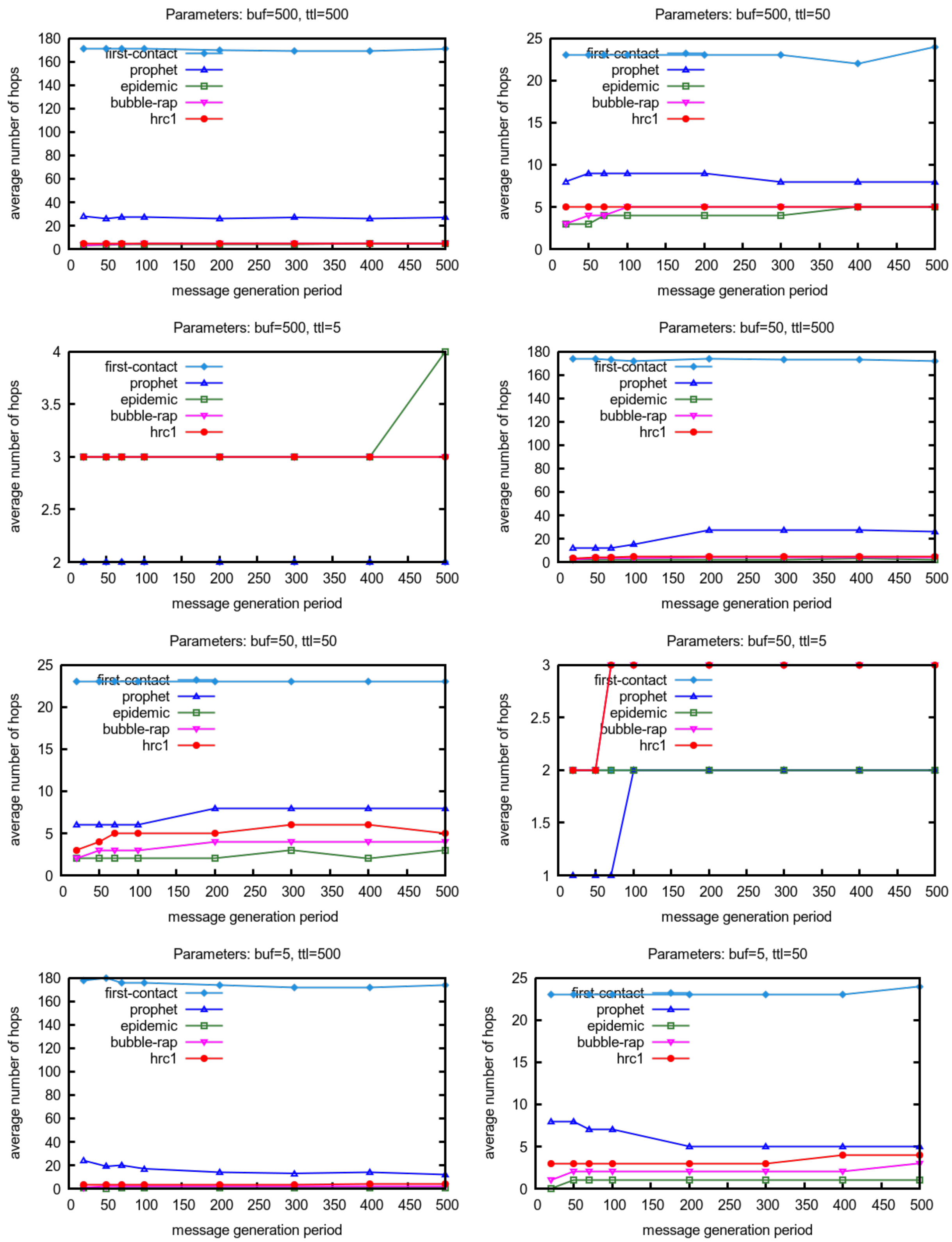

Figure 4.6: Average number of hops as a function of the period of message generation

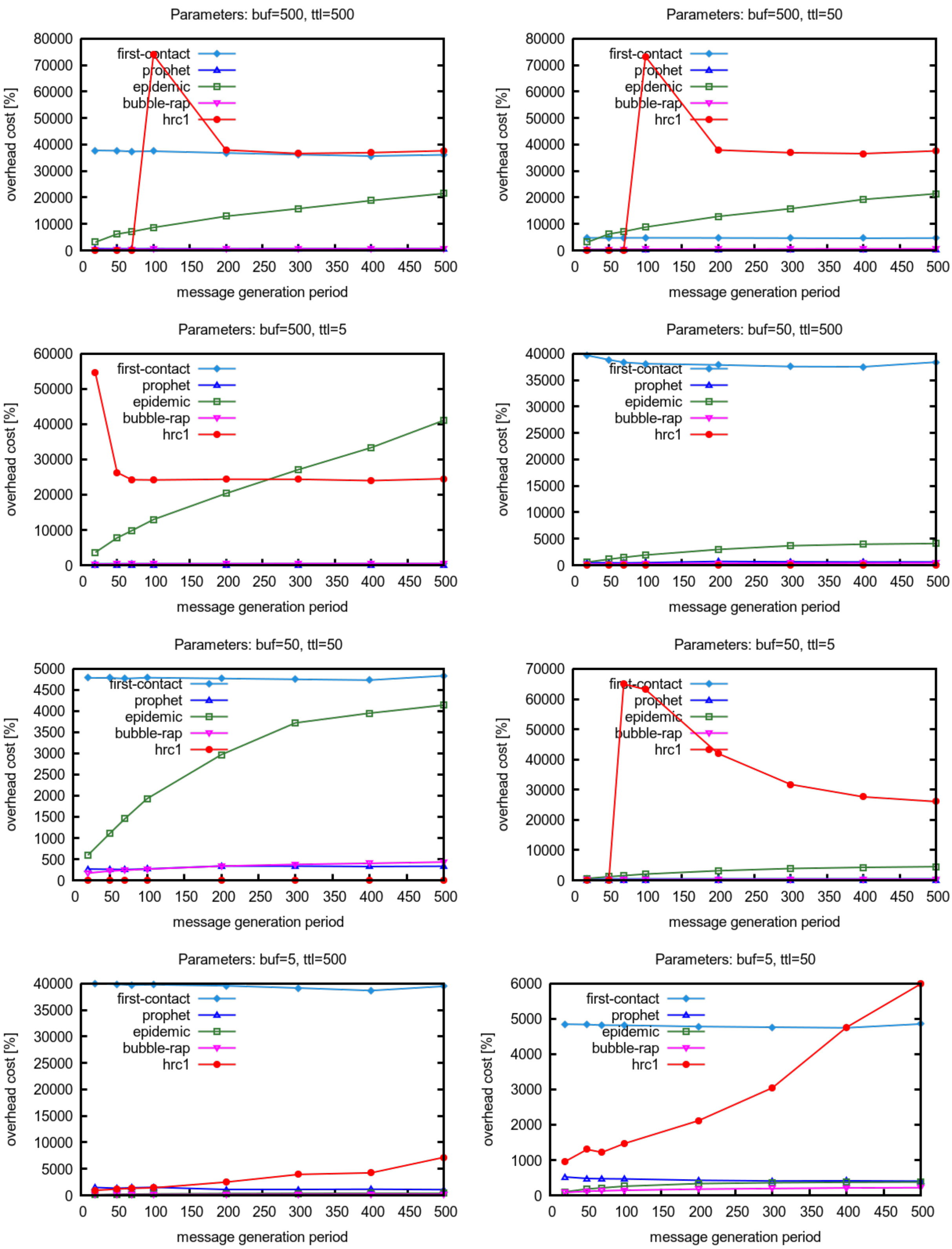

Figure 4.7: Overhead cost ratio as a function of the period of message generation

4. set of simulations: TTL = 50, message generation period = 500

5. set of simulations: TTL = 50, message generation period = 100

6. set of simulations: TTL = 50, message generation period = 50

7. set of simulations: TTL = 5, message generation period = 500

8. set of simulations: TTL = 5, message generation period = 50

Both the TTL and message generation periods are referred in time units of simulation. 1 time unit = 2 s.

RESULTS:

Fig. 4.8 shows the message delivery ratio (referred as delivered messages [%] in graphs) as a function of the message buffer size. The proposed routing schema Hierarchical Routing with Clustering (HRC1) outperforms all other methods in the number of delivered messages. It reflects the fact that the routing method was designed primary with the respect to this characteristic. The best results have been achieved for combination of TTL = 500 time units (1000 s) and message generation period = 500 (1000 s), and for TTL = 50 time units (100 s) and message generation period = 500 (1000 s), where the message delivery rate ratio achieves almost 90 percent of delivered messages. It seems the main influence on message delivery ratio of the proposed method has the period of injection of new messages into system. For message generation period = 500 (1000 s) the message delivery ratio of HRC1 grows as a function of the side message buffer size until the buffer size equal to 200 messages and than stays flat. For shorter message generation periods is the message delivery ratio dependent on buffer size: the higher values of buffer size imply the higher values of message delivery ratio. The message delivery ratios of other methods have values between 10 and 20 percent of delivered messages and their graphs are almost flat.

Fig. 4.9 shows the average message delivery delay (referred as average delay in graphs) as a function of the message buffer size. Epidemic routing and First Contact routing outperform all other methods. This result is in accordance with our assumptions, both these methods use the simplest routing rules and routing is very fast. The average message delivery delay of the proposed routing schema HRC1 is comparable to BUBBLE-Rap for longer message generation periods. For message generation period = 50 (100 s) BUBBLE-Rap works slightly better. The average message delivery delay of the PROPHET protocol depends rather on TTL than on the size of node message buffer. For extremely low value of TTL, the PROPHET routing scheme achieves the worst results. The proposed routing scheme HRC1 achieves boldly high values of average message delivery delay for small sizes of message buffer. The average message delivery delay decreases with the increasing buffer size. The improvement in average message delivery delay, which appears with growing size of buffer, is high for small sizes of the buffer, but as the size of the buffer grows, the dependency on the buffer size becomes weak.

Fig. 4.10 shows the average number of hops as a function of the message buffer size. In accordance to our assumptions, highest values of this parameter are achieved by First

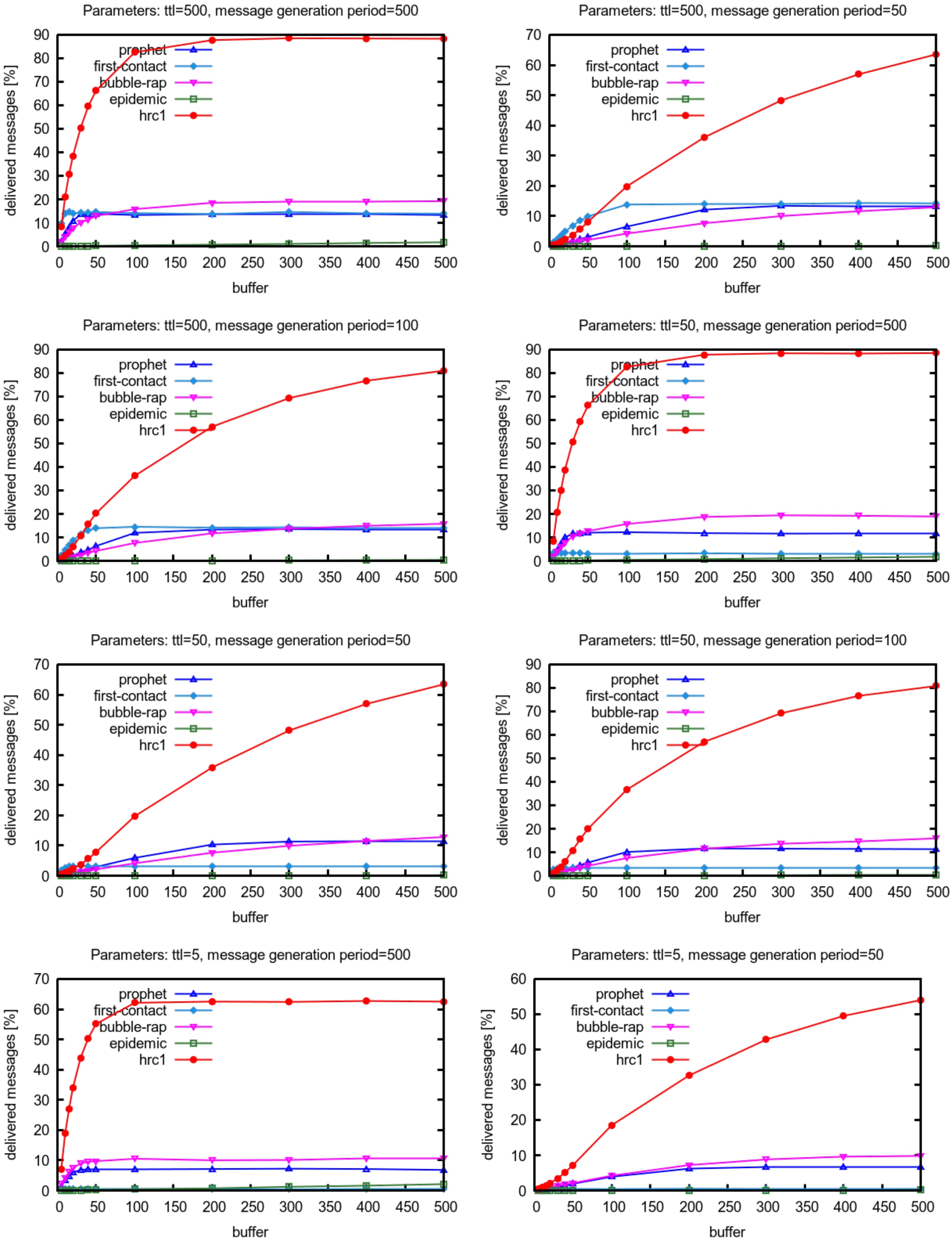

Figure 4.8: Message delivery ratio as a function of the size of node message buffer

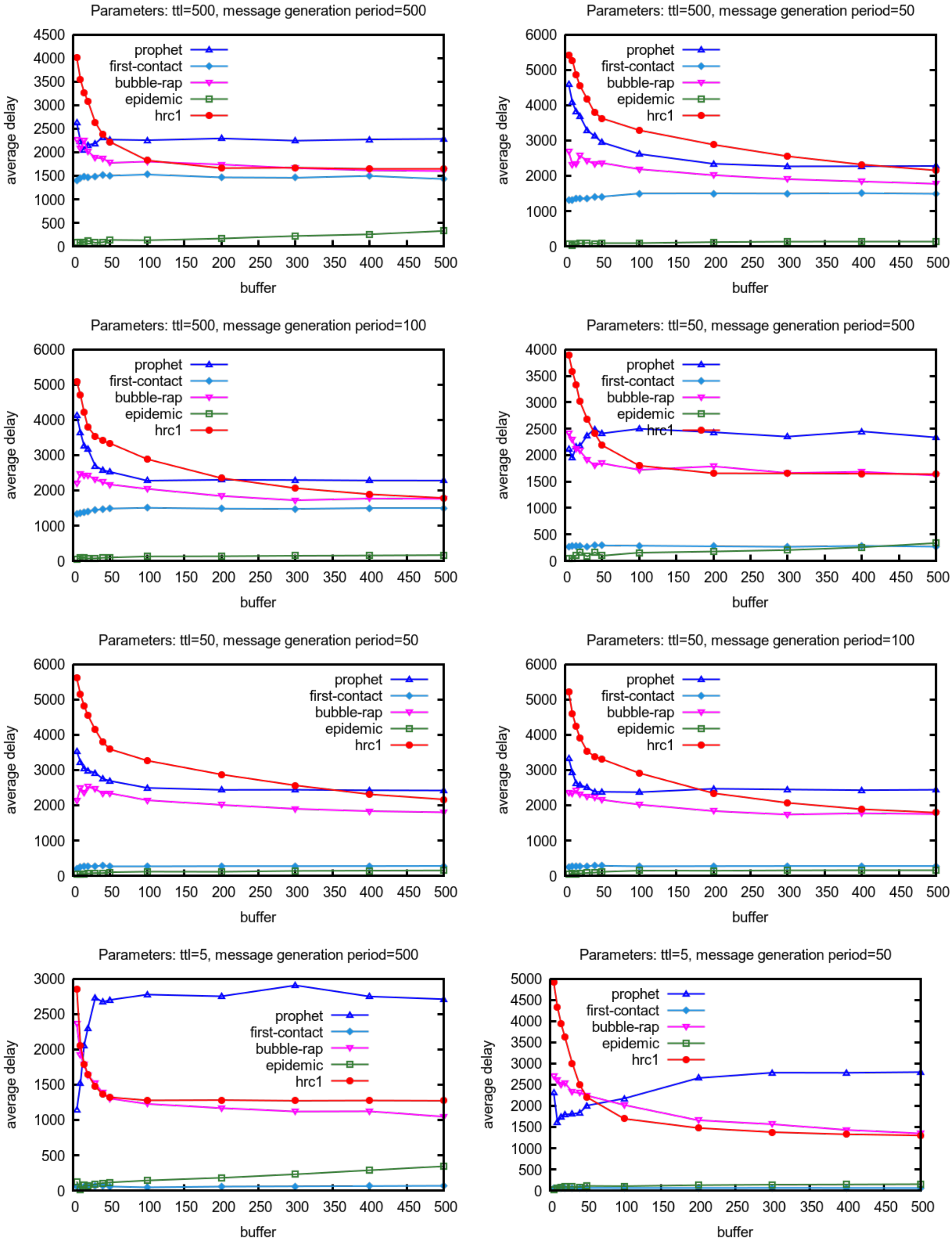

Figure 4.9: Message delivery delay as a function of the size of node message buffer

Contact routing scheme. The average number of hopes of the proposed routing schema HRC1 is comparable to BUBBLE-Rap and Epidemic Routing. The graphs are almost flat. It can be interpreted such as the number of hopes depends rather on routing method and message numeration period than on the size of node message buffer.

Fig. 4.11 shows the overhead cost ratio (referred as overhead cost in graphs) as function of the message buffer size. The graphs indicates that the proposed method has the worst results in overhead ratio, but it works without network congestion. The lover values of overhead of Epidemic routing are caused by the metrics which we use for overhead computation. Generated, but never set messages are not taken into account in this metrics. We can observe unexpected extremely high peaks of overhead cost ratio for HRC1 routing scheme for several combinations of the values of TTL and message generation period and buffer size 100 or 400. The configuration with these parameters leads to computation of large communities. HRC1 routing scheme uses the approach, that the routing inside the communities is epidemic routing with predefined timeout. The occurrences of peaks of overhead ratio can be eliminated by implementation of more tight constraints of application of epidemic routing into method or by limiting the number of nodes forming communities.

4.4.1.4 Influence of Time-to-live (TTL) to Performance Metrics

We conducted simulations for different values of TTL (from 1 to 300 time units). 1 simulation time unit is equal to 2 s. We observed the influence of TTL to the performance metrics. The other simulation parameters the size of node message buffer and message generation period (referred as send period in graphs) were set as follows:

1. set of simulations: buffer = 500 messages, message generation period = 500
2. set of simulations: buffer = 500 messages, message generation period = 100
3. set of simulations: buffer = 500 messages, message generation period = 50
4. set of simulations: buffer = 50 messages, message generation period = 500
5. set of simulations: buffer = 50 messages, message generation period = 100
6. set of simulations: buffer = 50 messages, message generation period = 50
7. set of simulations: buffer = 5 messages, message generation period = 500
8. set of simulations: buffer = 5 messages, message generation period = 50

Both the TTL and message generation periods are referred in time units of simulation. 1 time unit = 2 s.

Influence of TTL to message delivery ratio Fig. 4.12 shows the message delivery ratio (referred as delivered messages [%] in graphs) as function of TTL. The proposed routing schema Hierarchical Routing with Clustering (HRC1) outperforms all other methods in the number of delivered messages. The best results have been achieved for combination

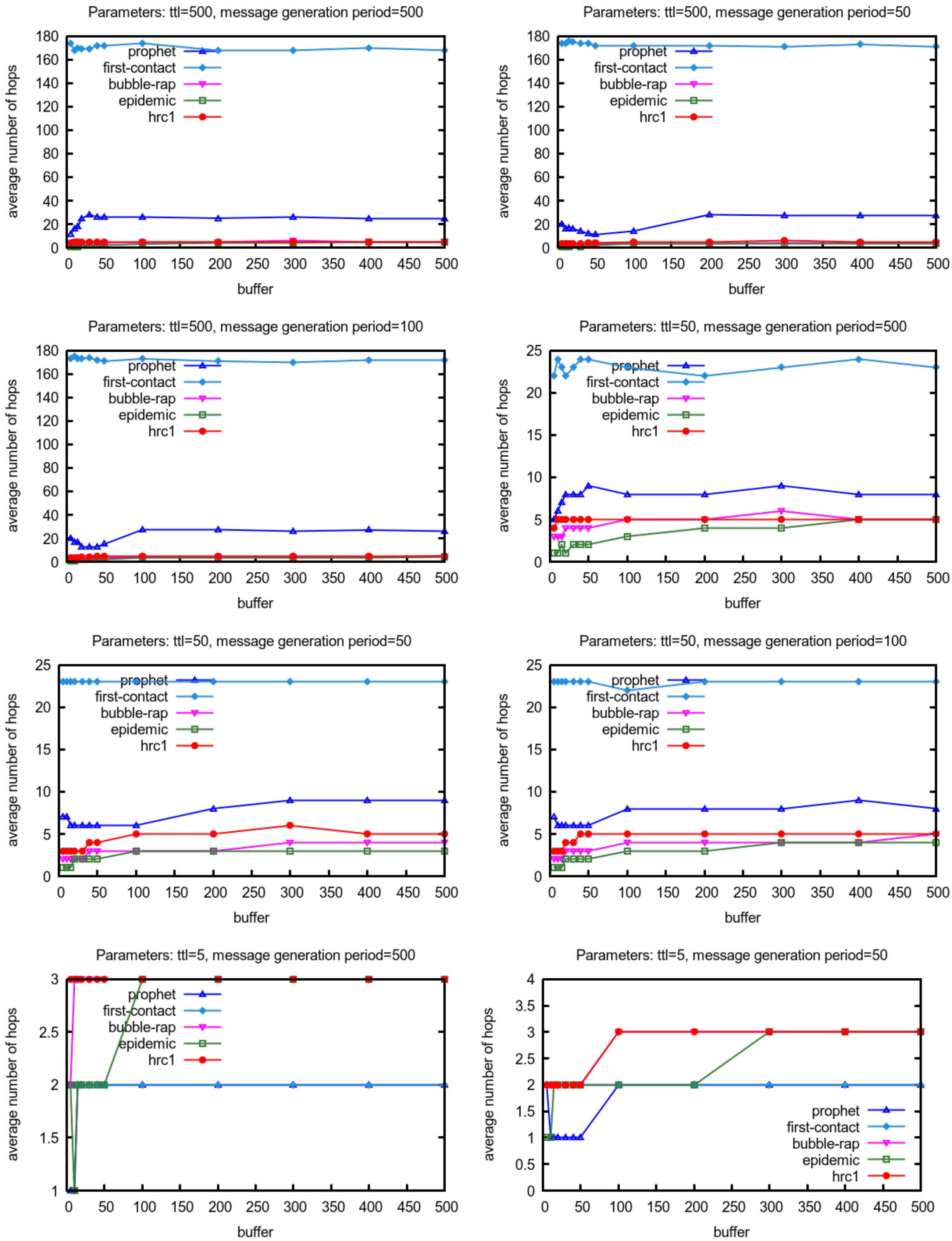

Figure 4.10: Average number of hops as a function of the size of node message buffer

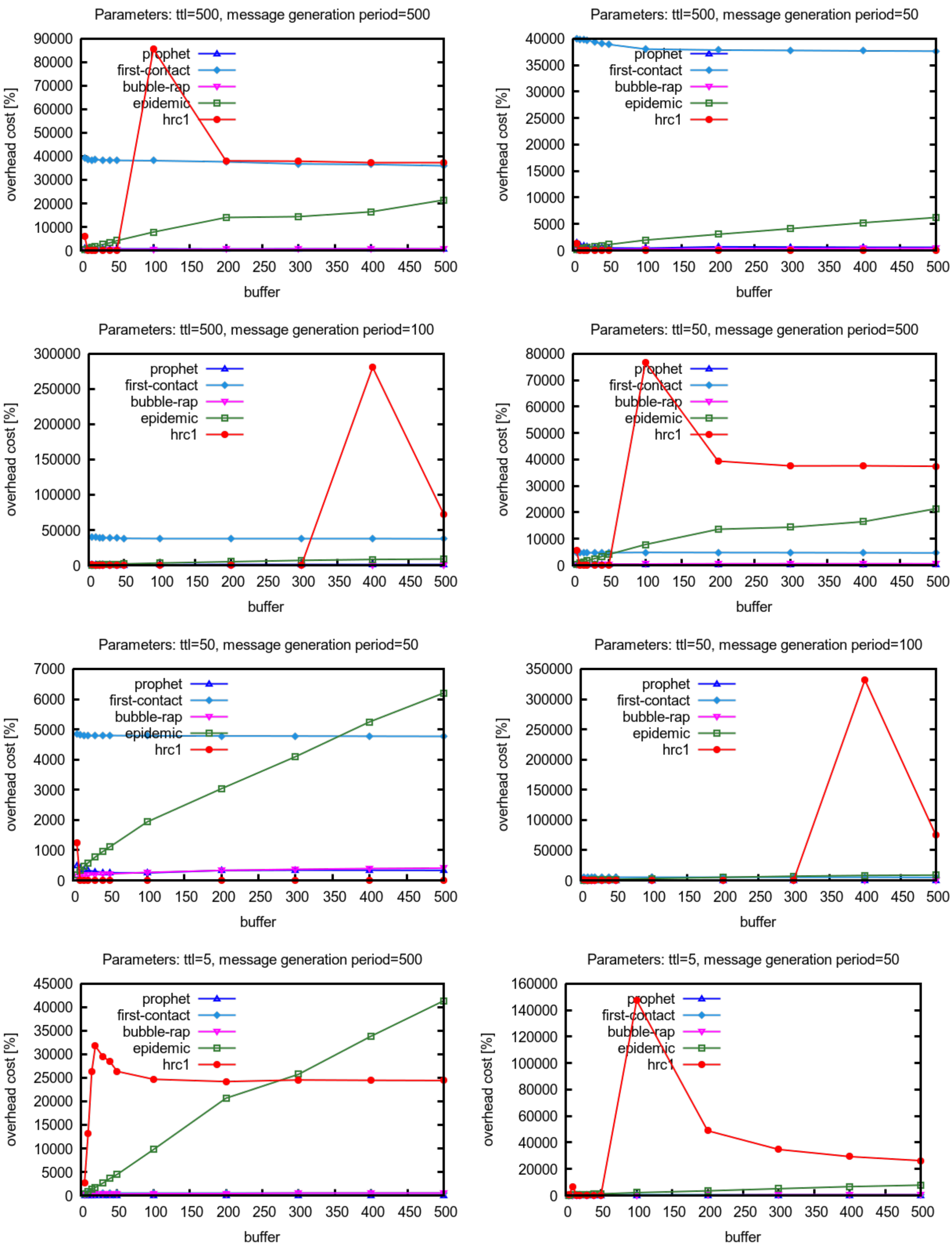

Figure 4.11: Overhead cost ratio as a function of the size of node message buffer

of buffer size = 500 messages and message generation period = 500 (1000 s). For this combination of parameters, about nightly percent of generated messages were delivered. A very good message delivery ratio can be achieved also for combination and for TTL = 50 time units (100 s) and message generation period = 500 (1000 s), where the message delivery rate ratio achieves almost 90 percent of delivered messages. It seems the main influence on message delivery ratio of the proposed method has the period of injection of new messages into system. For message generation period = 500 (1000 s) the message delivery ratio of HRC1 grows as a function of the side message buffer size until the buffer size equal to 200 messages and than stays flat. For shorter message generation periods is the message delivery ratio dependent on buffer size: the higher values of buffer size imply the higher values of message delivery ratio. The message delivery ratios of other methods have values between 10 and 20 percent of delivered messages and their graphs are almost flat. It can be interpreted as follows: in OPNs, where node moves in accordance with some complex temporal pattern, the message delivery ratio is influenced rather by the accuracy of the node position prediction strategy than by the size of node message buffer.

Fig. 4.13 shows the average message delivery delay (referred as average delay in graphs) as the function of TTL. The linear dependency of the average message delivery delay is observable for the First Contact routing scheme. Other wise, the graphs indicates low values of average message delivery delay for extremely small values of TTL, but most proposed methods have poor performance in message delivery pro small values of TTL. It implies that these low message delivery delays for extremely small values of TTL are probably caused by the overall low number of delivered messages in the system.

Fig. 4.14 shows the average number of hops as function of average number of hops. The graphs indicate strong linear dependence between the number of the hopes and TTL for the First Contact routing scheme and small dependence between the number of the hopes and TTL for the PROPHET routing scheme, otherwise the graphs are almost flat. The average number of hopes of the proposed routing schema HRC2 is low and it is comparable to Epidemic routing and BUBBLE-Rap. It can be interpreted as considering that the number of hopes doesn't not depend on TTL for the implemented routing method of HRC1, BUBBLE-Rap or epidemic routing and that there is a strong linear dependency on TTL for the First Contact routing scheme.

Fig. 4.15 shows the overhead cost ratio (referred as overhead cost in graphs) as function of message buffer size. shows overhead cost ratio as a function of TTL. The graphs indicates no import correlation between TTL and proposed method HRC2, except the correlations observable for extremely small values of TTL. For he extremely small values of TTL, HRC2 does not perform well on this simulation scenario.

4.4.2 Experiment 2

Simulation Scenario 2: Geographically structured OPN (Venice)
Our tested method: Svm-based routing (svm)
Compared to: Epidemic Routing, PROPHET, First Contact, Bubble Rap

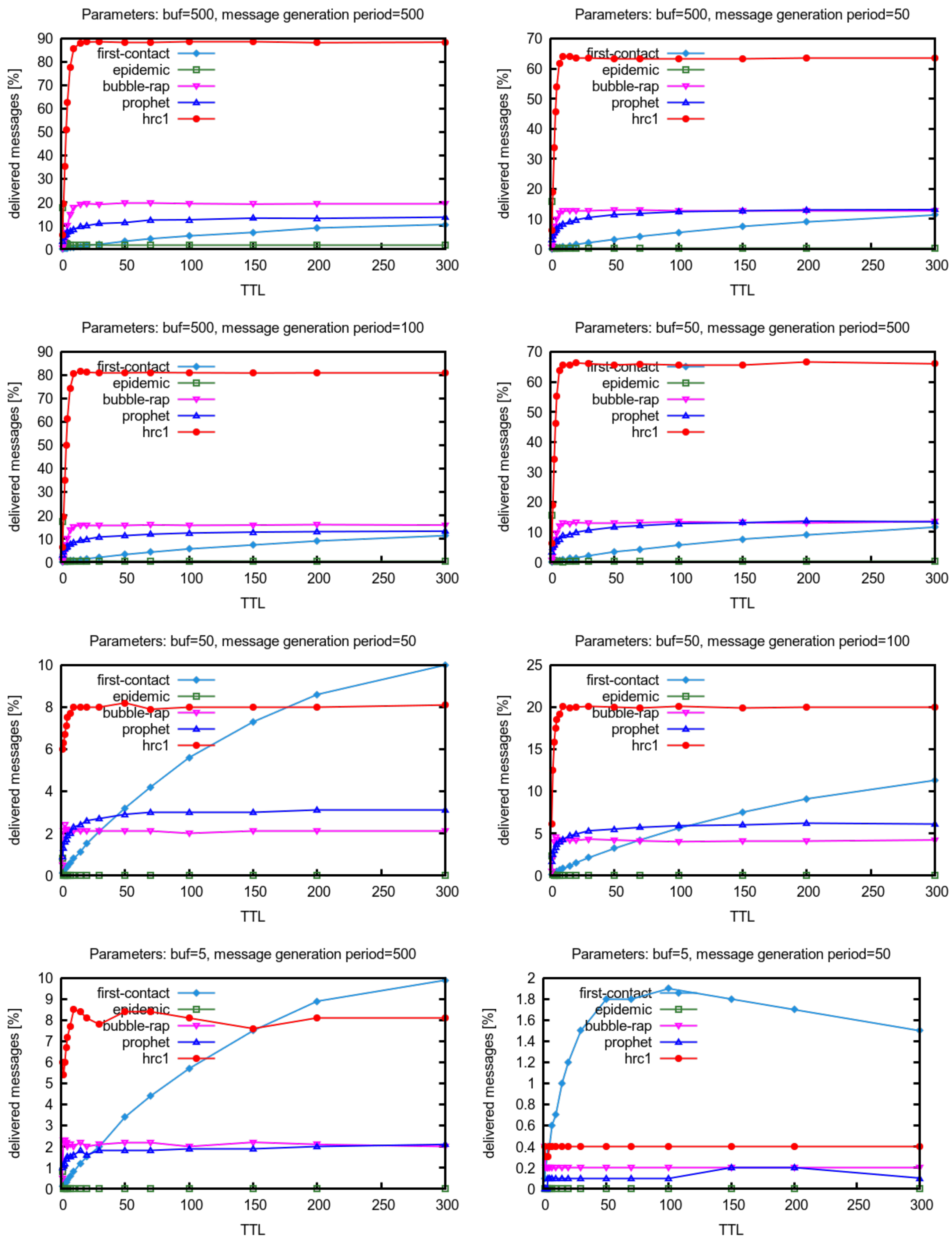

Figure 4.12: Message delivery ratio as a function of the time to live

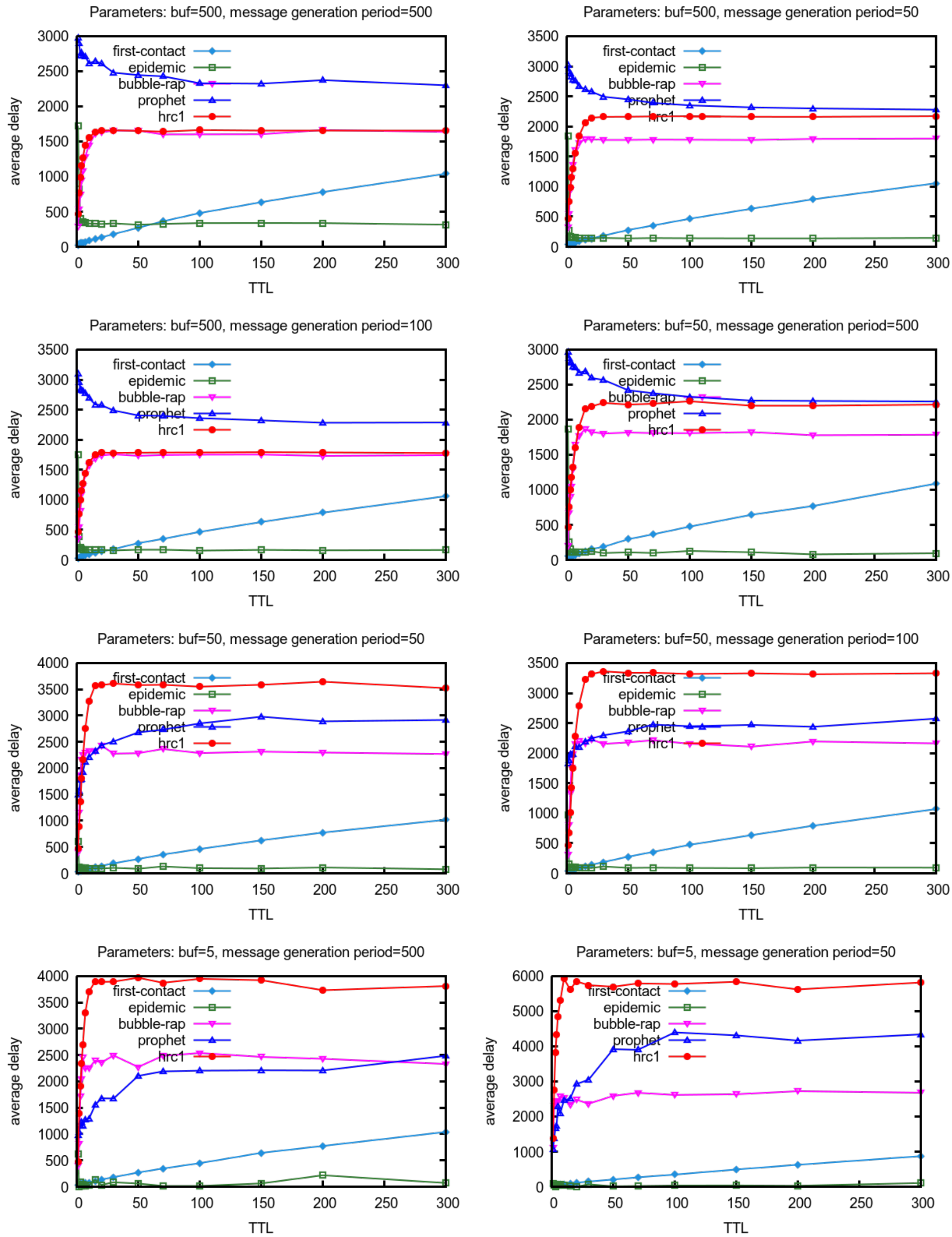

Figure 4.13: Message delivery delay as a function of the time to live

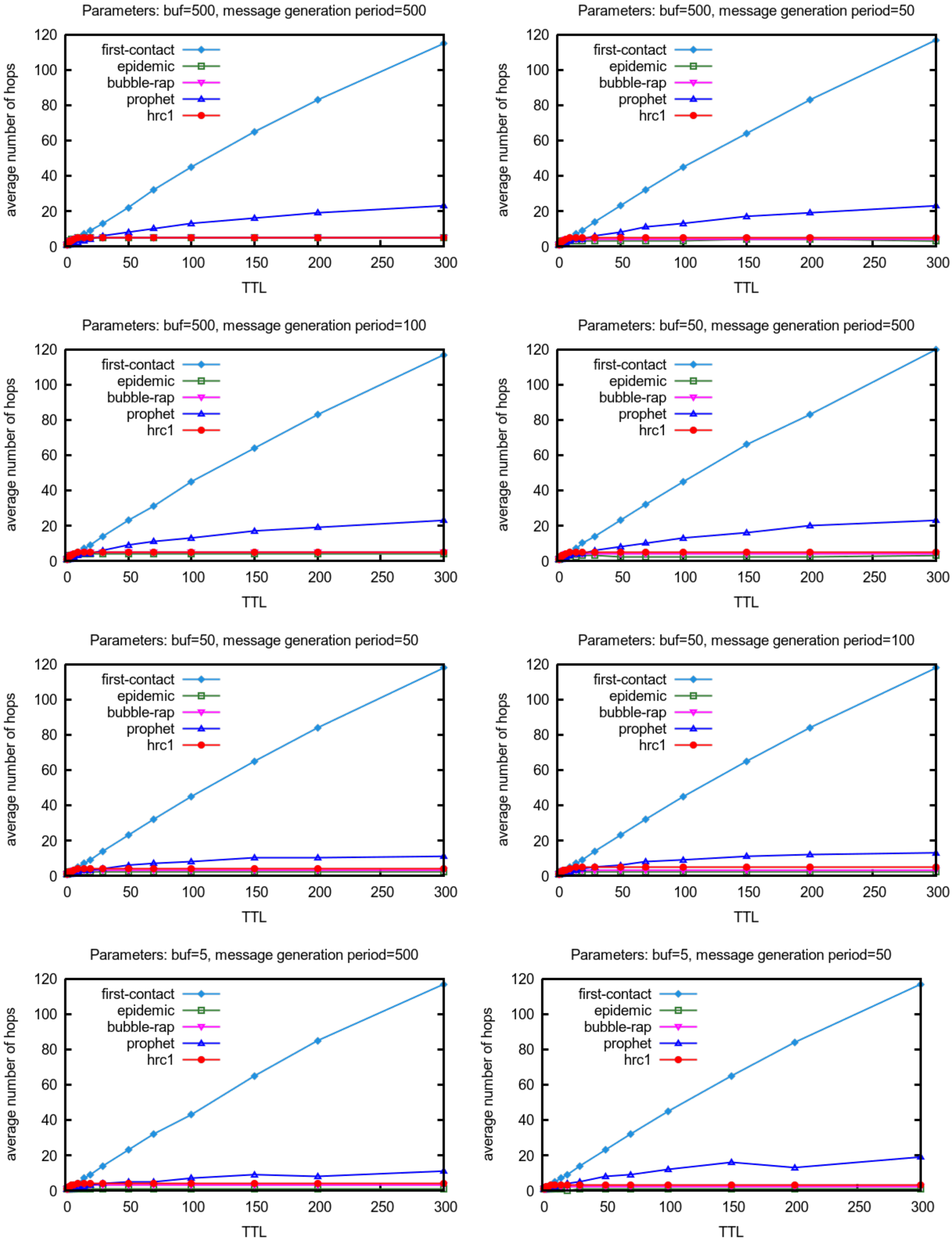

Figure 4.14: Average number of hops as a function of the time to live

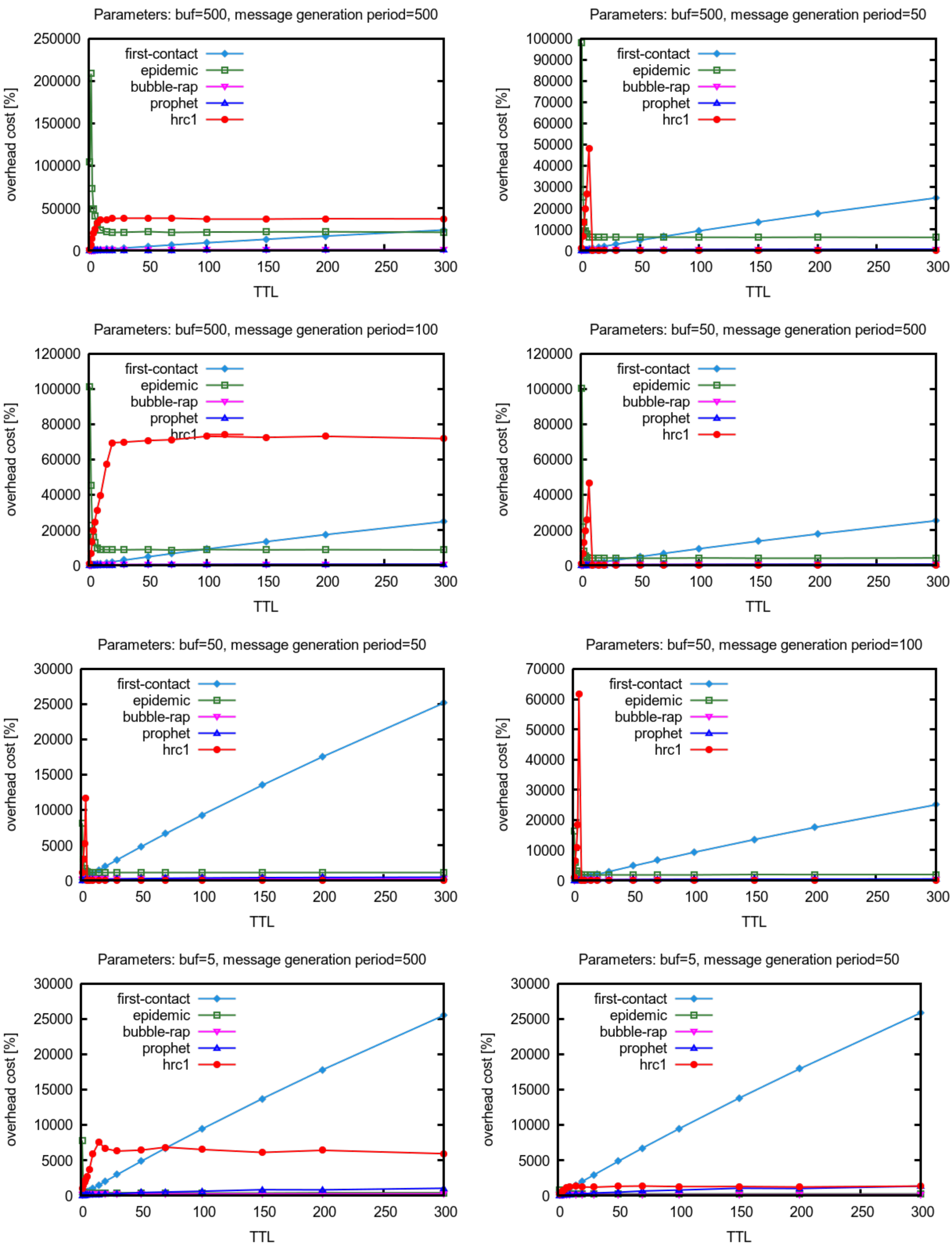

Figure 4.15: Overhead cost ratio as a function of the the time to live

Main characteristics: Two geographical regions
Changing number of nodes visiting both regions
All nodes generate messages in predefined time period

Parameter	Range
Map	Venice
Simulation size	4500 x 3400 m
Moving speed	random 0.5 - 1.5 m/s
Transmission range	50 m
Simulation Time	432000 time units ~ 10 days
Sampling Period T_s	0.5 time units ~ 1 second
Message size	36 bytes
Node Buffer Size	1 - 500 messages
Message Generation Period	100 time units ~ 200 seconds
Time to live	50 transmissions

Table 4.2: Simulation Scenario 1: Simulation setup for ONE Simulator

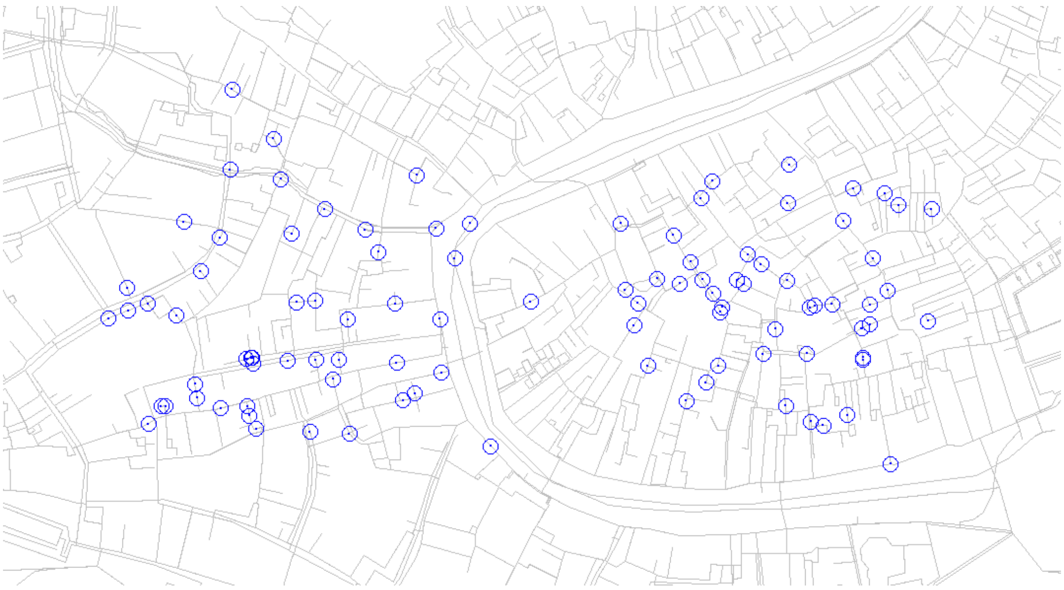

Figure 4.16: Simulation Scenario 2: The initial positions of nodes in OPN geographical area

Fig. 4.16 shows the initial positions of nodes in OPN geographical area.

This experiment was conducted on simulation scenario 2. The OPN consists of two separated target regions located in the urban area of Venice. The selected geographical area is characterized by the high density of roads and by the water channel which can be crossed by nodes only at some places (bridges). We selected the uniform probabilistic distribution of targets in each region. We selected the uniform probabilistic distribution of node initial positions in each OPN geographical region and log-normal probabilistic distribution of node stays in targets during the day phase. The simulation was conducted for 10-day cycle with periodically changing day traffic pattern. For the purposes of OPN routing, we selected the data collected in time interval of 3 hours (7 AM to 10 AM) from each day of the 10-day simulation. The analyzed time interval was limited to 3 hours per day primary for the computational reasons. The selected 3-hour interval consists of mobility patterns changing each twenty minutes. The data from four days were used to train the model. The data from six days were used to test the performance of the proposed method. The number of nodes in simulation was 100 nodes. Fig. 4.17 shows the simulation results.

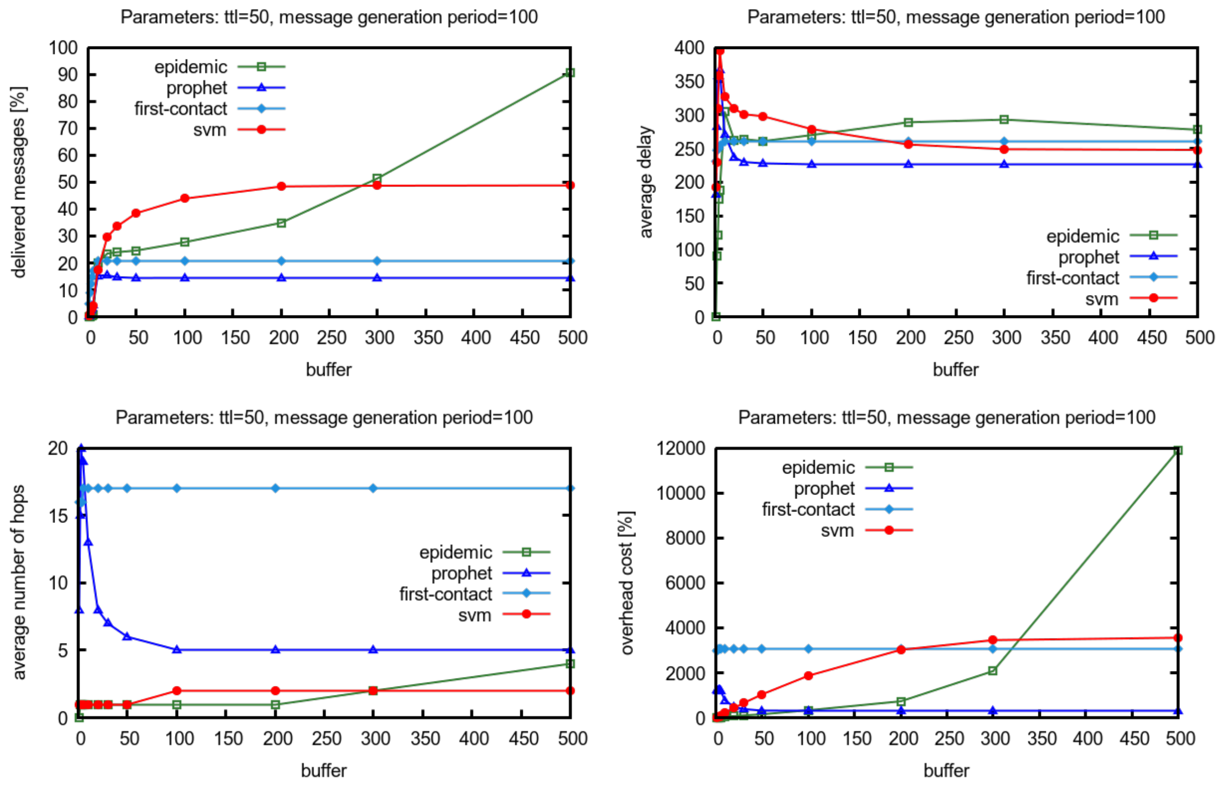

Figure 4.17: SVM-based routing results on Simulation Scenario 2: Message delivery ratio, average message delays, average number of hops and overhead cost ratio as a function of the period of the size of node message buffer

RESULTS: Fig. 4.17 hows the message delivery ratio (referred as delivered messages [ratio as functions of the buffer size in simulation. The graphs indicate, that all compared methods achieve approximately the same message delivery delay for the simulations with buffer=¿100 messages. For the low values of buffer size, the average message delivery delay increases. This is in accordance to the routing rules: when the node message buffer is full, the node is not able to store more messages, and the delivery delay increases. The average number of hopes of the proposed method for this simulation is low, about 2-3 hopes. The highest value of average hopes is achieved by First Contact routing protocol. The graph indicates, that the proposed method SVM-based routing has a good overhead. For the combination of parameters TTL=50 time units (100s) and message generation period=100 time units (200s), it can be observed that the overhead cost ratio of SVM-based routing protocol does not depend on the size of buffer since the buffer size cross the threshold 300 messages. For lower values, an almost linear dependency is observable. The proposed method outperforms other tested methods in message delivery ratio, except Epidemic routing for scenarios with higher values of buffer size.

4.4.3 Experiment 3

Scenario: Simulation Scenario 3 : Regions in Line
Our tested method: Hierarchical routing with clustering 2 (HRC2)
Compared to: Epidemic Routing, PROPHET
All nodes generate messages in predefined time period

Main characteristics:
Four regions
Two kinds of nodes: nodes visiting targets in one region, nodes visiting targets in two regions; there is no node visiting more than two regions
Undirect connection among regions (Fig. 4.19)

Parameter	Range
Map	Venice
Simulation size	4500 x 3400 m
Moving speed	random 0.5 - 1.5 m/s
Transmission range	20 m
Simulation Time	432000 time units ~ 10 days
Sampling Period T_s	0.5 time units ~ 1 second
Message size	36 bytes
Node Buffer Size	1 - 500 messages
Message Generation Period	10 - 500 time units ~ 20 - 1000 seconds
Time to live	1 - 300 transmissions

Table 4.3: Simulation Scenario 3: Simulation setup for ONE Simulator

Fig. 4.18 shows the initial positions of nodes in OPN geographical area.

Figure 4.18: Simulation Scenario 3: The initial positions of nodes in OPN geographical area

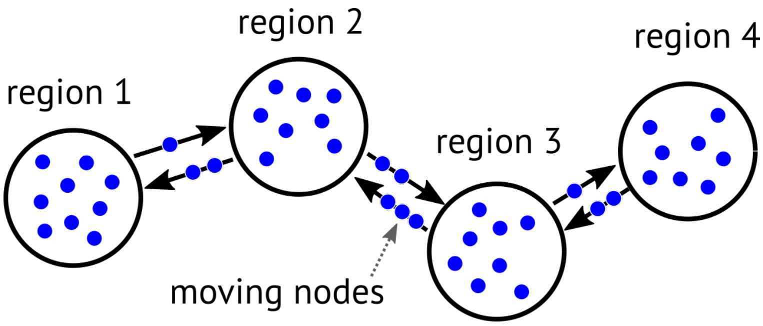

Figure 4.19: The schematic structure of the connection of OPN geographic regions in Simulation Scenario 3

4.4.3.1 Influence of Message Generation Period to Performance Metrics

Simulation Setup We conducted simulations for different of message generation period (referred as send period in graphs) and observed the influence of different message generation period to the performance metrics. The range of message generation periods was from 1 to 500. Lower values of message generation period imply higher rates of message generation by nodes, and consequently higher number of messages, which are simultaneously present in simulation. The other simulation parameters TTL and node message buffer were set as follows:

1. set of simulations: TTL = 500, buffer = 500

2. set of simulations: TTL = 500, buffer = 50

3. set of simulations: TTL = 500, buffer = 50
4. set of simulations: TTL = 50, buffer = 500
5. set of simulations: TTL = 50, buffer = 100
6. set of simulations: TTL = 50, buffer = 50
7. set of simulations: TTL = 5, buffer = 500
8. set of simulations: TTL = 5, buffer = 50

Fig. 4.20 shows the message delivery ratio (referred as delivered messages [%] in graphs) as a function of the period of message generation. The range of observed message generation period was an interval of 1 to 500. Lower values of message generation period imply higher rates of message generation by nodes, and consequently higher number of messages, which are simultaneously present in simulation. As it is shown in graphs, the extremely low values of message generation period cause network congestion and the message delivery ratio rapidly decreases. Furthermore, it can be observed that the achieved results strongly depend on the size of message buffer. For the extremely short buffers (buffer = 5), the performance of the proposed method decreases and the message delivery ratio of HRC1 is about one percent. The best results were achieved for the large buffer (buffer = 500). The proposed routing schema HRC1 outperforms all other methods in the number of delivered messages for all tested combinations of parameters TTL and buffer. The best results have been achieved for the combination of TTL = 500 time units (1000 s) and the buffer size = 500, and for the combination of TTL = 50 time units (100 s) and the buffer size = 500, where the message delivery rate ratio achieves almost 60 percent of delivered messages (respectively 25 percent) on this simulation scenario. It seems that the highest influence on message delivery ratio has the combination of all three parameters: buffer size, TTL and message generation period. The appropriate size of message buffer is crucial. The method does not perform well for extremely low values of message buffer. We observe two types of dependencies between the message generation period and message delivery ratio. For the large values of the buffer parameter, the graph indicates rather logarithmic increase in message delivery ratio. For the small values of the buffer parameter, the graph indicates rather linear increase in message delivery ratio.

Fig. 4.21 shows the average message delivery delay (referred as average delay in graphs) as a function of the message generation period. The graph indicates high values of average message delivery delay for extremely small values of the message generation period. As the period increases, the graph becomes flat. For the simulation scenarios with buffer=500, the average message delivery delay of the proposed method HRC2 is comparable to this one of PROPHET. The average message delivery delay of the epidemic routing is the best.

Fig. 4.22 shows the average number of hops as a function of the message generation period. In accordance to our assumptions, the high number of hopes can be observed when the PROPHET routing scheme was in use in simulation. The average number of hopes of the proposed routing schema HRC2 is low and it is comparable to Epidemic routing. The

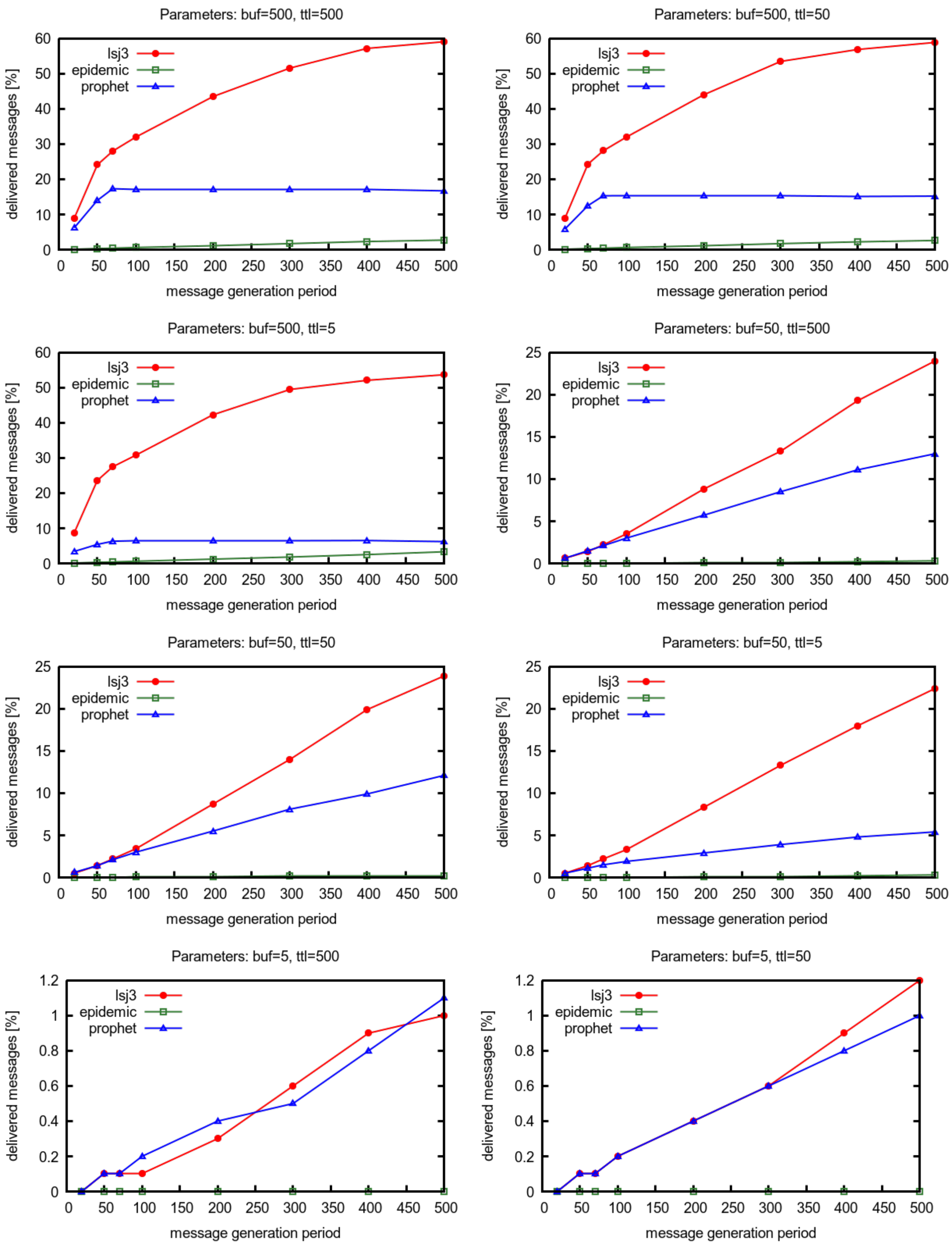

Figure 4.20: Simulation Scenario 4: Message delivery ratio as a function of the period of message generation

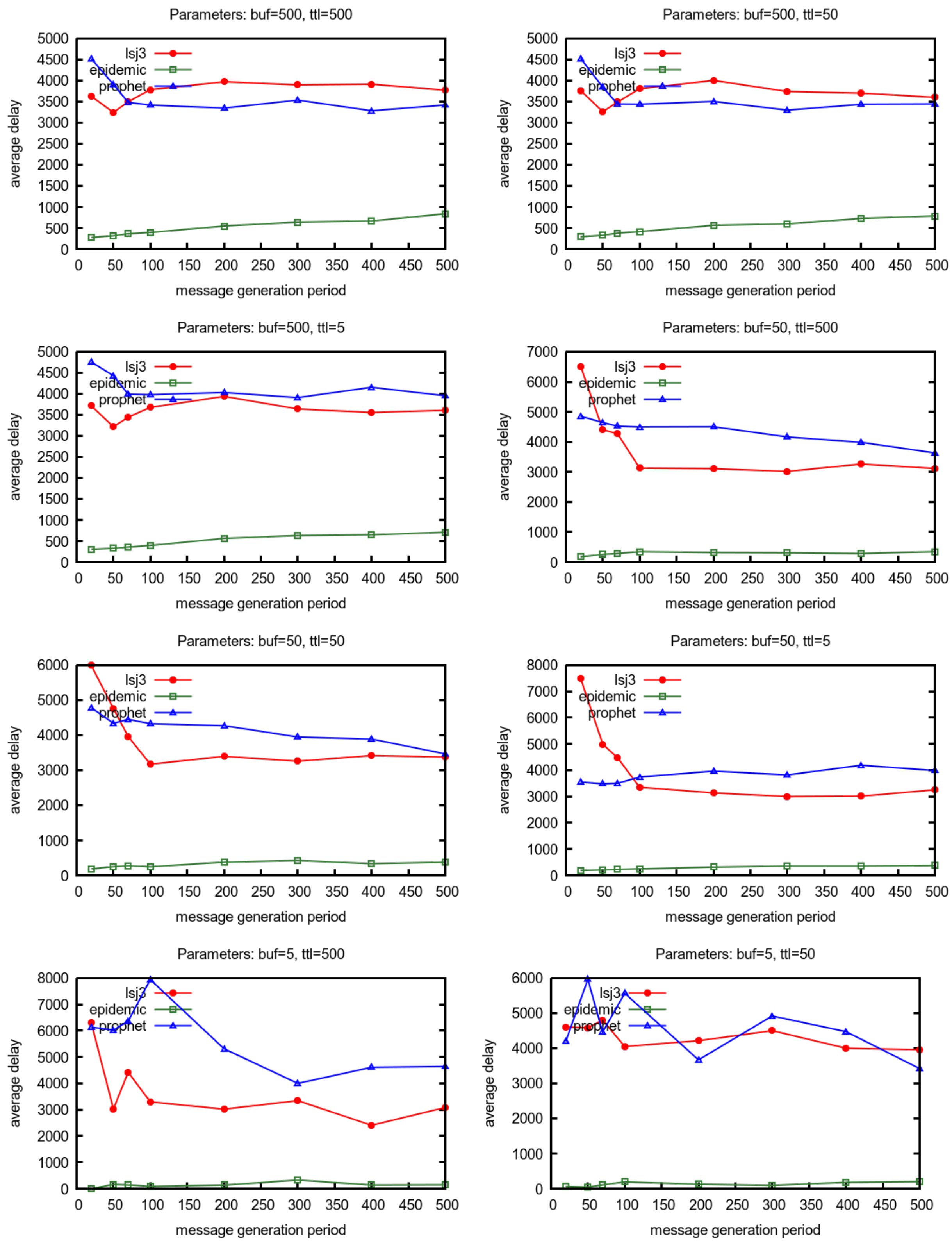

Figure 4.21: Simulation Scenario 3: Message delivery delay as a function of the period of message generation

graphs for HRC2 Epidemic routing are almost flat. It can be interpreted as considering that the number of hopes depends particularly on the applied routing method.

Influence of message generation period size to overhead cost ratio

Fig. 4.23 shows the overhead cost ratio (referred as overhead cost in graphs) as the function of message generation period. shows the overhead cost ratio (referred as overhead cost in graphs) as the function of the message generation period. The graphs indicates that the proposed method has high overhead. The lover values of overhead of Epidemic routing are caused by the metrics which we use for overhead computation. Generated, but never set messages are not taken into account. We can observe unexpected extremely high peaks of overhead cost ratio for HRC2 routing scheme for several combinations of the values of TTL and message generation period and buffer size of 300 or 400. The configuration with these parameters leads to computation of large communities. HRC2 routing scheme uses the approach, that the routing inside the communities is epidemic routing with timeout. The occurrences of peaks of overhead ratio can be eliminated by implementation of more tight constraints of application of epidemic routing into method or by limiting the number of nodes forming communities.

4.4.3.2 Influence of Buffer Size to Performance Metrics

We conducted simulations for simulation for different sizes of message buffer (from 1 to 500 messages) and observed the influence of message buffer size to the performance metrics. The other simulation parameters TTL and message generation period (referred as send period in graphs) were set as follows:

1. set of simulations: TTL = 500, message generation period = 500
2. set of simulations: TTL = 500, message generation period = 100
3. set of simulations: TTL = 500, message generation period = 50
4. set of simulations: TTL = 50, message generation period = 500
5. set of simulations: TTL = 50, message generation period = 100
6. set of simulations: TTL = 50, message generation period = 50
7. set of simulations: TTL = 5, message generation period = 500
8. set of simulations: TTL = 5, message generation period = 50

Both the TTL and message generation periods are referred in time units of simulation. 1 time unit = 2 s.

Influence of node message buffer size to message delivery ratio Fig. 4.24 shows the message delivery ratio (referred as delivered messages [%] in graphs) as a function of the message buffer size. The proposed routing schema HRC2 outperforms all other methods in the number of delivered messages. The graph suggests the linear dependence between

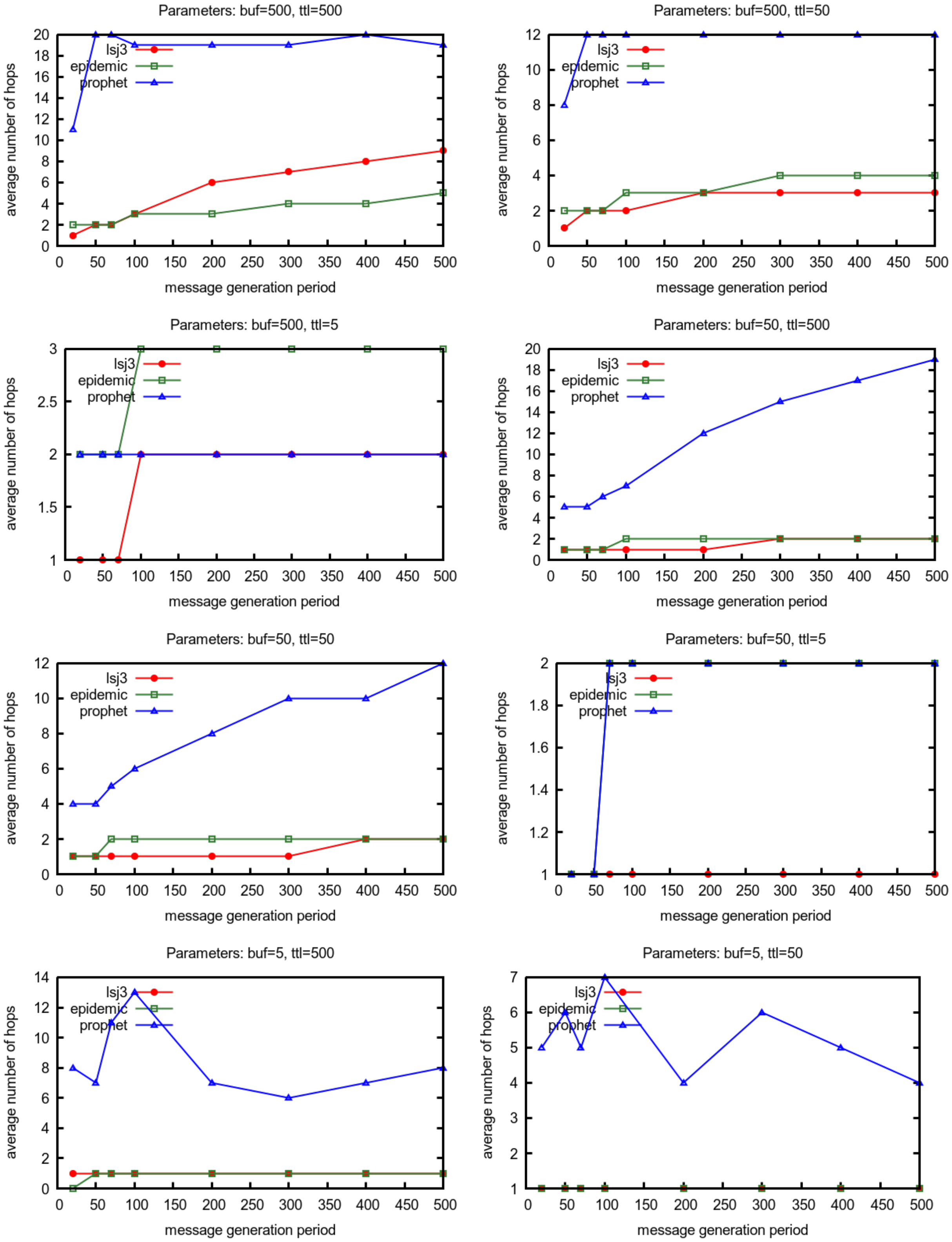

Figure 4.22: Simulation Scenario 3: Average number of hops as a function of the period of message generation

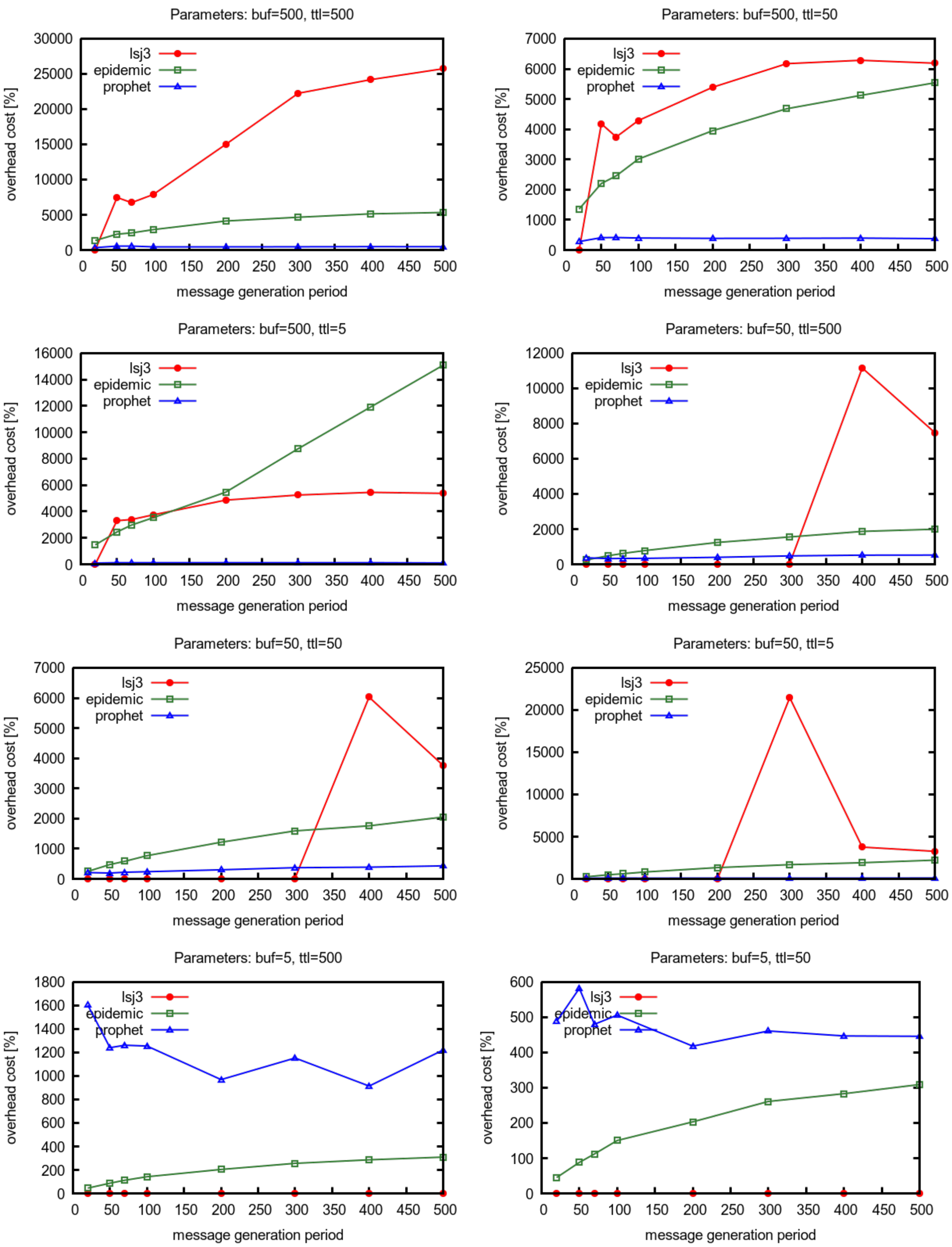

Figure 4.23: Simulation Scenario 3: Overhead cost ratio as a function of the period of message generation

buffer size and message delivery ratio for lower values of buffer. As the value of buffer increases the graphs become more flat. We observe the influence of the all three simulation parameters TTL, message generation period and TTL to method performance. The best results have been achieved for combination of TTL = 500 or 50 time units (1000s or 100s) and message generation period = 500 (1000 s), about sixty percent.

Influence of node message buffer size to message delivery delay Fig. 4.24 shows the average message delivery delay (referred as average delay [%] in graphs) as the function of the message buffer size. The graph suggests the the high influence of the buffer size for low values of buffer between buffer size and message delivery ratio for lower values of buffer. As the value of buffer increases the graphs become more flat. Since the buffer size is 100, no significant dependencies on message delivery delay to buffer size can be observed. The proposed methods HRC2 performance is comparable to PROPHET. Epidemic routing outperforms both the HRC2 and PROPHET.

Influence of node message buffer size to number of hopes Fig. 4.24 shows the average number of hops (referred as average delay [%] in graphs) as the function of the message buffer size. The graph suggests the the proposed methods HRC2 performance is comparable to those one of Epidemic routing in comparison to PROPHET where the average number of hopes reaches 20 for the scenarios, where TTL=500 (1000s), respectively 12 for the scenarios where TTL was set to lower numbers.

Fig. 4.27 shows the overhead cost ratio (referred as overhead cost in graphs) as the function of the message buffer size. The graphs indicates that the proposed method is the worst of the compared methods form the view of point of overhead, but it works without network congestion. The lover values of overhead of Epidemic routing are caused by the metrics which we use for overhead computation. Generated, but never set messages are not taken into account. We can observe unexpected extremely high peaks of overhead cost ratio for HRC2 routing scheme for several combinations of the values of TTL and message generation period and buffer size 200 or 400. The configuration with these parameters leads to computation of large communities. HRC2 routing scheme uses the approach, that the routing inside the communities is epidemic routing with timeout. The occurrences of peaks of overhead ratio can be eliminated by implementation of more tight constraints of application of epidemic routing into method or by limiting the number of nodes forming communities. Both the TTL and message generation periods are referred in time units of simulation. 1 time unit = 2 s.

4.4.3.3 Influence of Time-to-live (TTL) to Performance Metrics

We conducted simulations for different values of TTL (from 1 to 300 time units). 1 simulation time unit is equal to 2 s. We observed the influence of TTL to the performance metrics. The other simulation parameters the size of node message buffer and message generation period (referred as send period in graphs) were set as follows:

1. set of simulations: buffer = 500 messages, message generation period = 500

2. set of simulations: buffer = 500 messages, message generation period = 100

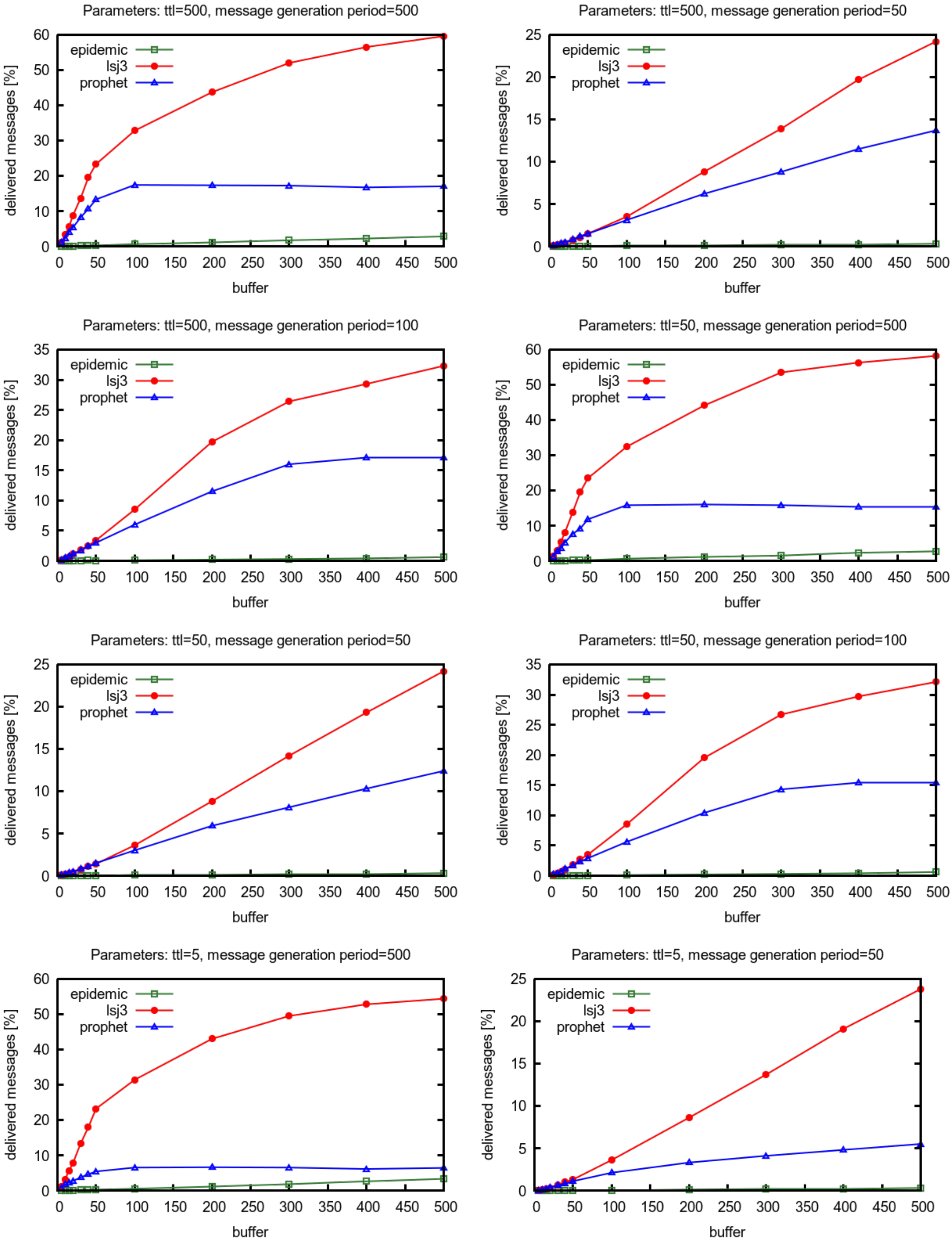

Figure 4.24: Simulation Scenario 3: Message delivery ratio as a function of the size of node message buffer

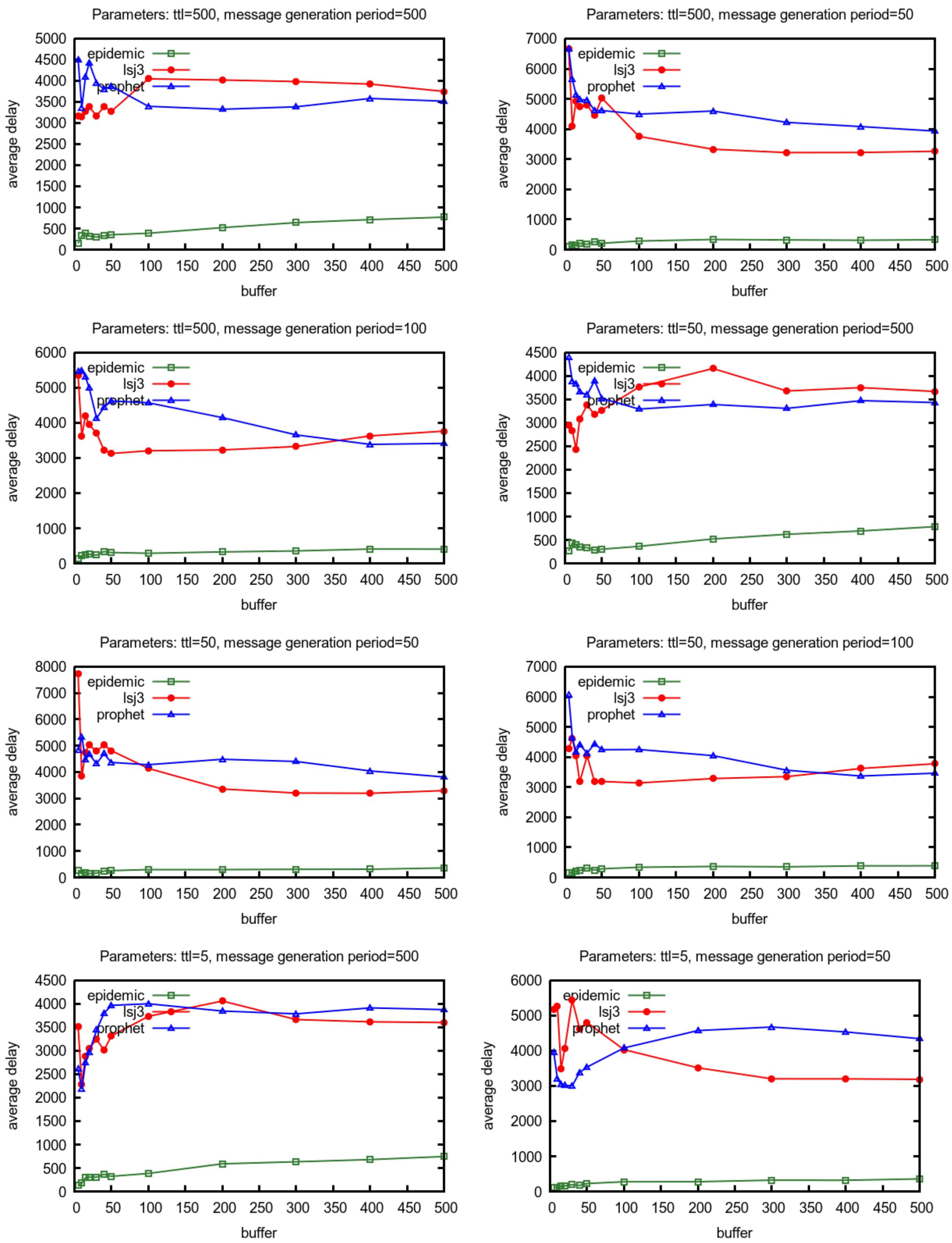

Figure 4.25: Simulation Scenario 3: Message delivery delay as a function of the size of node message buffer

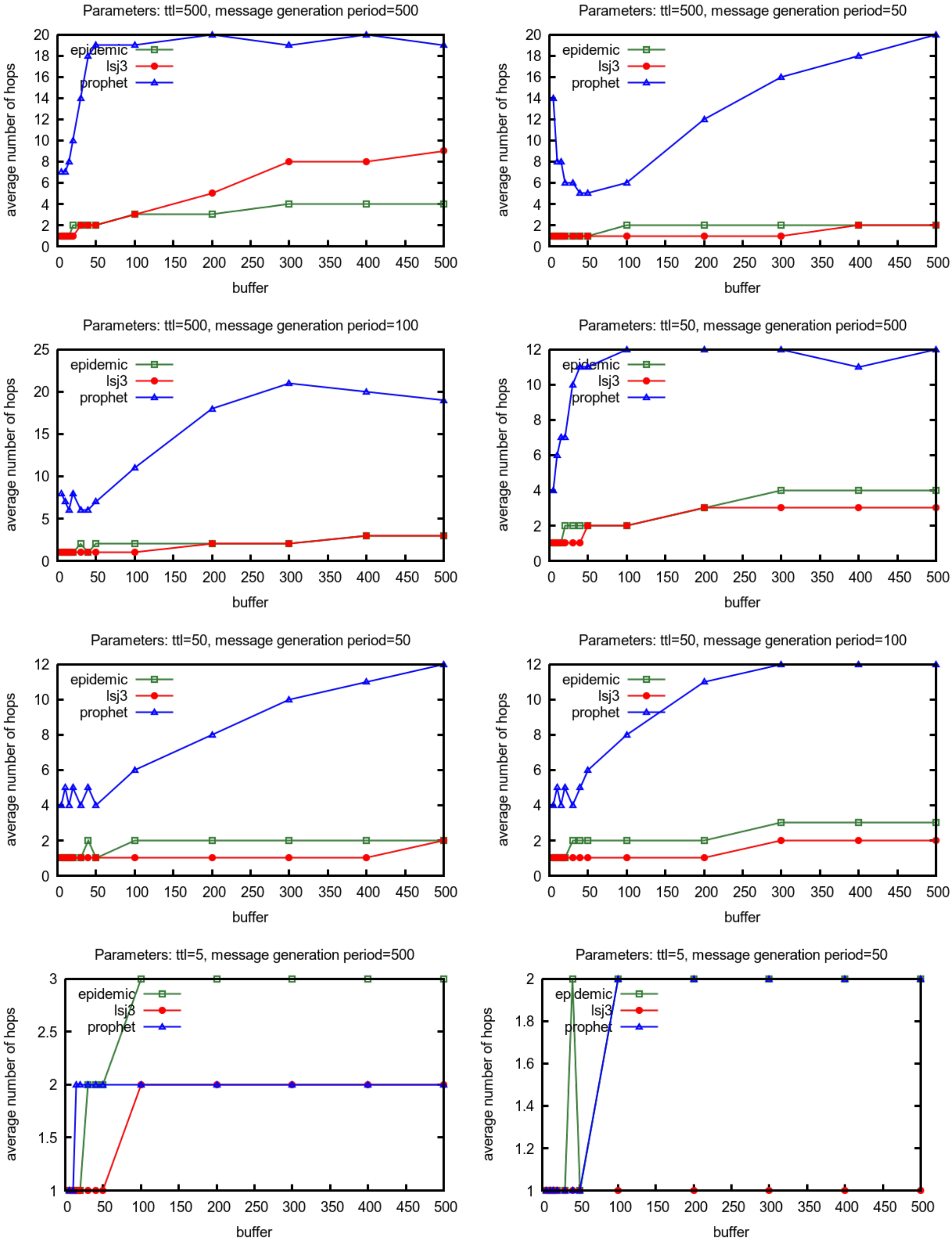

Figure 4.26: Simulation Scenario 3: Average number of hops as a function of the size of node message buffer

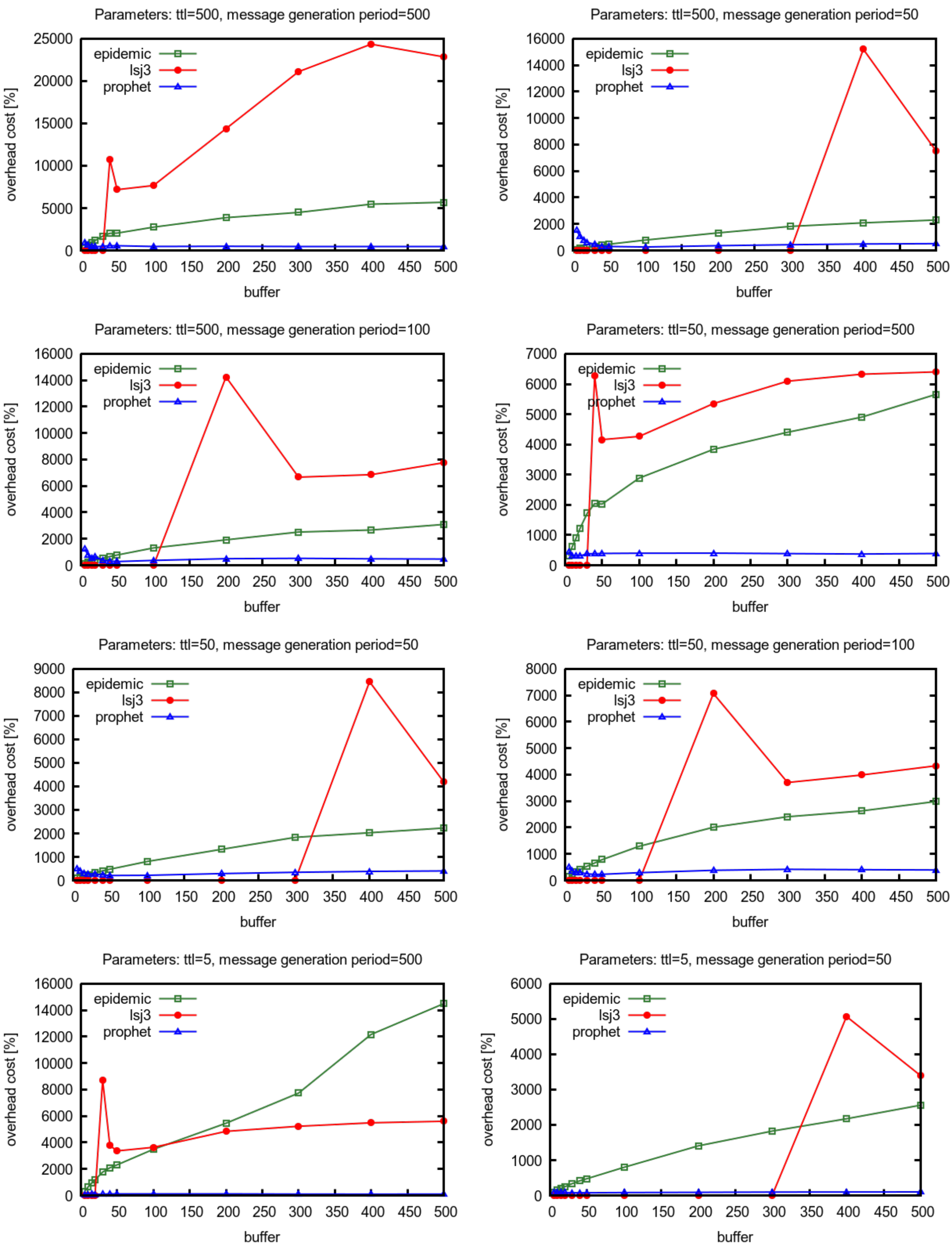

Figure 4.27: Simulation Scenario 3: Overhead cost ratio as a function of the size of node message buffer

3. set of simulations: buffer = 500 messages, message generation period = 50
4. set of simulations: buffer = 50 messages, message generation period = 500
5. set of simulations: buffer = 50 messages, message generation period = 100
6. set of simulations: buffer = 50 messages, message generation period = 50
7. set of simulations: buffer = 5 messages, message generation period = 500
8. set of simulations: buffer = 5 messages, message generation period = 50

Both the TTL and message generation periods are referred in time units of simulation. 1 time unit = 2 s.

Influence of TTL to message delivery ratio Fig. 4.28 shows the message delivery ratio (referred as delivered messages [%] in graphs) as function of TTL. Except the graph parts related to small values of TTL, no significant dependencies between TTL and average message delivery delay can be observed. For small values of TTL, the performance of all compared methods decreases and the message delivery delay increases.

Fig. 4.29 shows the average message delivery delay (referred as average delay in graphs) as the function of TTL. The graphs are flat. The graphs indicate low values of average message delivery delay for the small values of TTL, but the most analyzed methods have poor performance in message delivery for small values of TTL. It implies that these low message delivery delays for the small values of TTL are probably caused by the overall low number of delivered messages in the system.

Fig. 4.30 shows the average number of hops as function of TTL. The graphs indicate logarithmic dependence between the number of the hopes and TTL for the PROPHET routing scheme, otherwise the graphs are almost flat. The average number of hopes of the proposed routing schema HRC2 is low and it is comparable to Epidemic routing. It can be interpreted as considering that the number of hopes does not depend on TTL for the implemented routing method of HRC2 and Epidemic routing for this kind of scenario, but there is a logarithmic dependency on TTL for the PROPHET in this particular simulation scenario.

Fig. 4.31 shows overhead cost ratio as a function of TTL. The graphs indicates no import correlation between TTL and proposed method HRC2, except the correlations observable for extremely small values of TTL.

4.4.4 Experiment 4

Simulation Scenario 4: Random (Helsinki)
Our tested method: Hierarchical routing with clustering 2 (hrc 2)
Compared to: Epidemic Routing, PROPHET, First Contact, Bubble Rap
All nodes generate messages in predefined time period

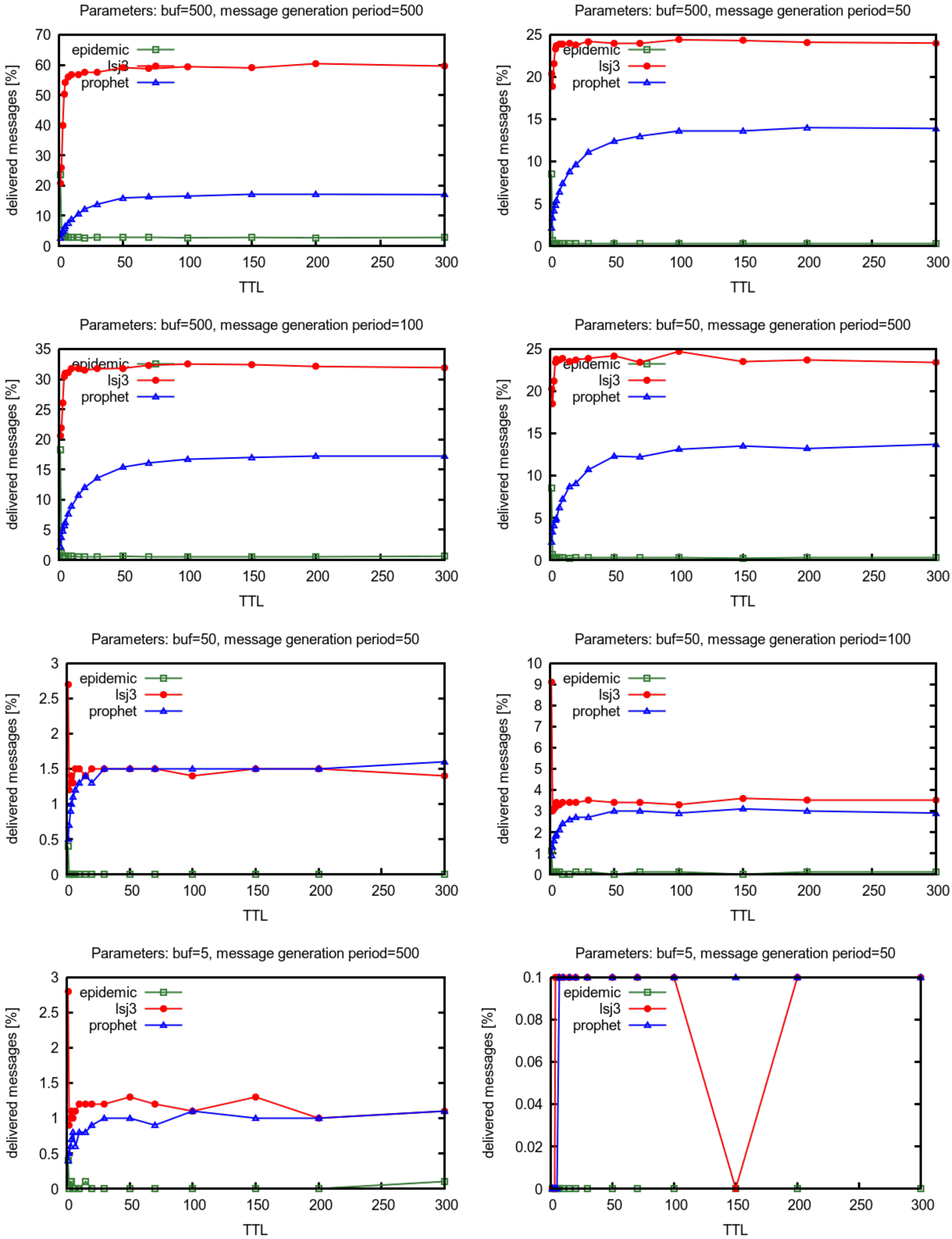

Figure 4.28: Simulation Scenario 3: Message delivery ratio as a function of the time to live

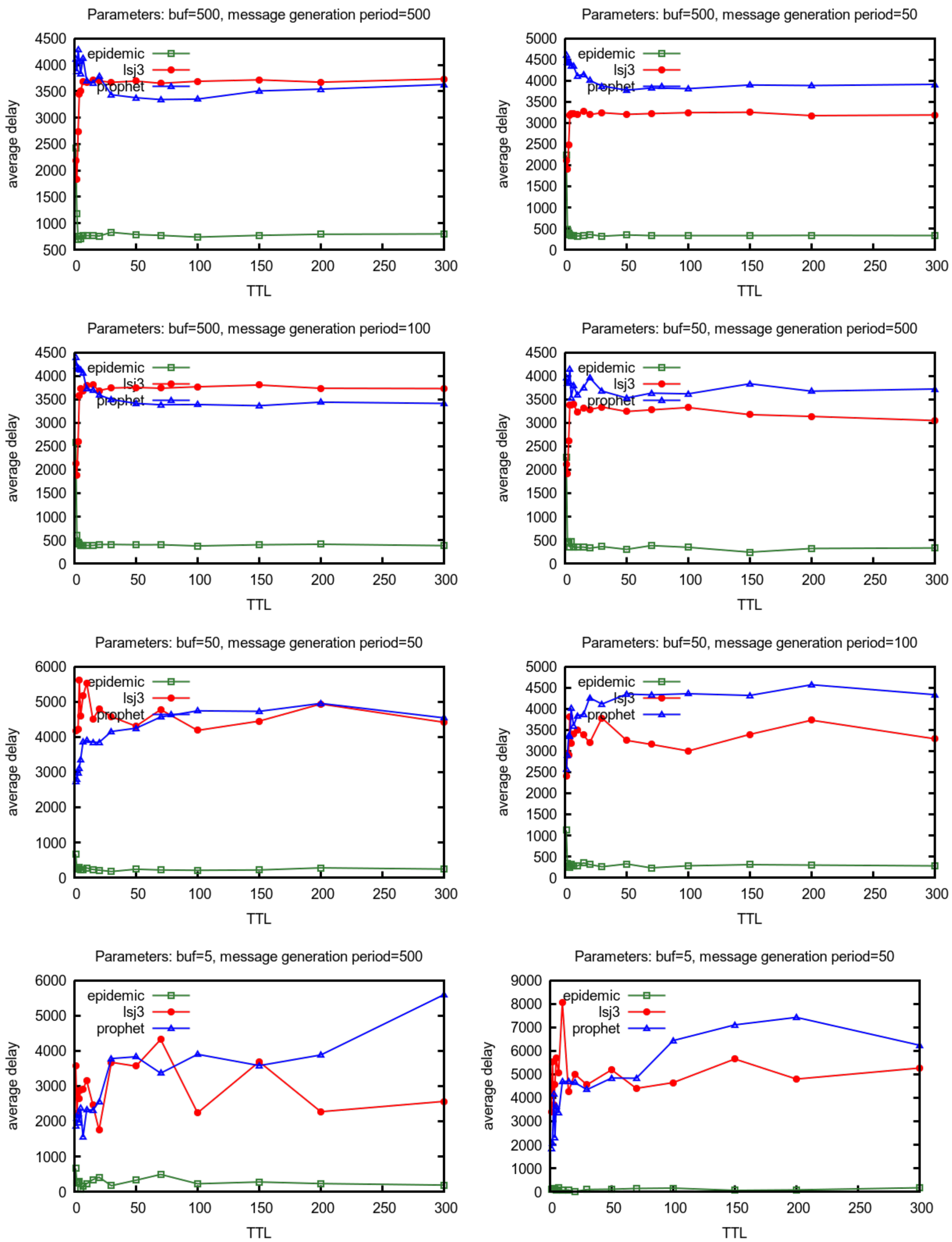

Figure 4.29: Simulation Scenario 3: Message delivery delay as a function of the time to live

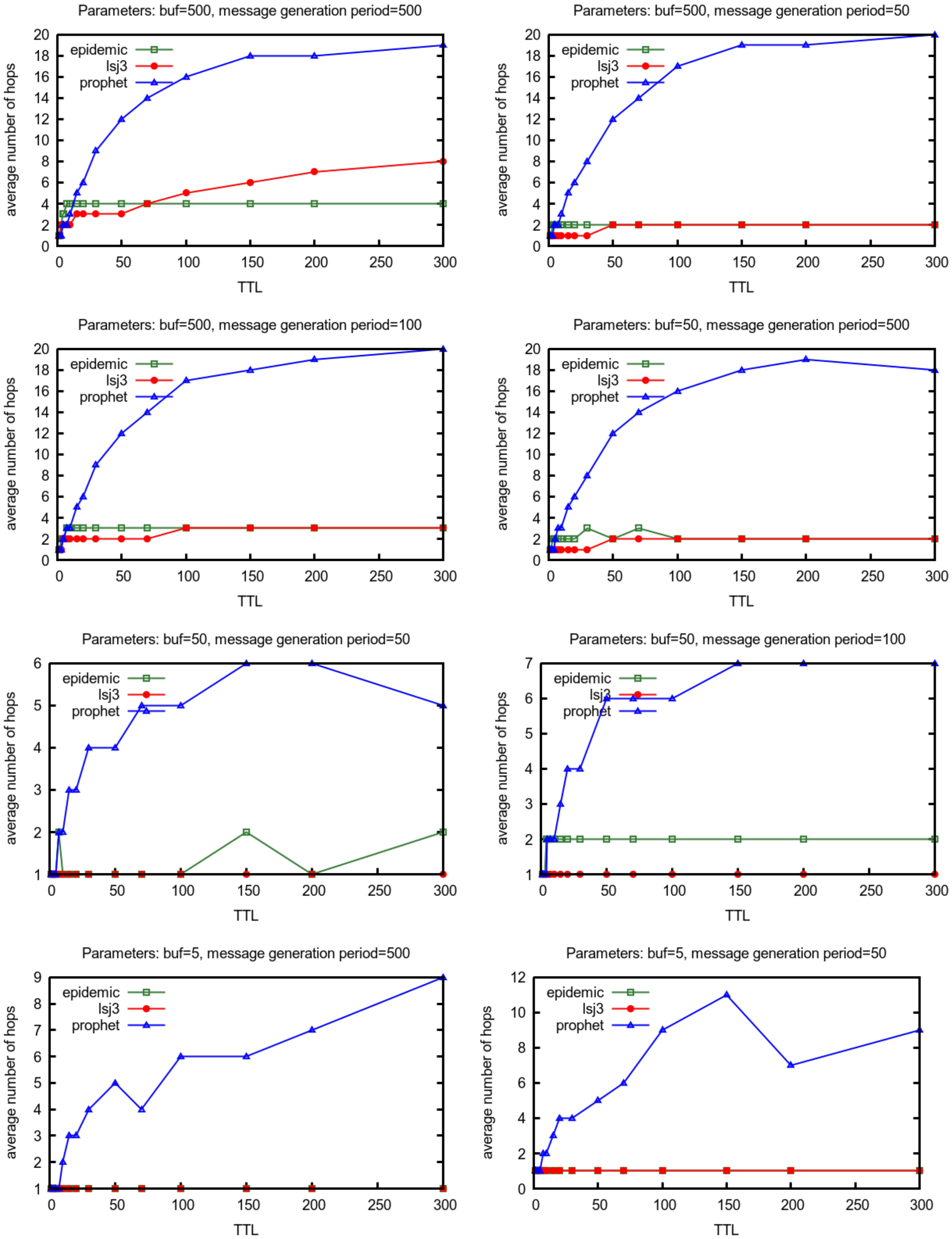

Figure 4.30: Simulation Scenario 3: Average number of hops as a function of the time to live

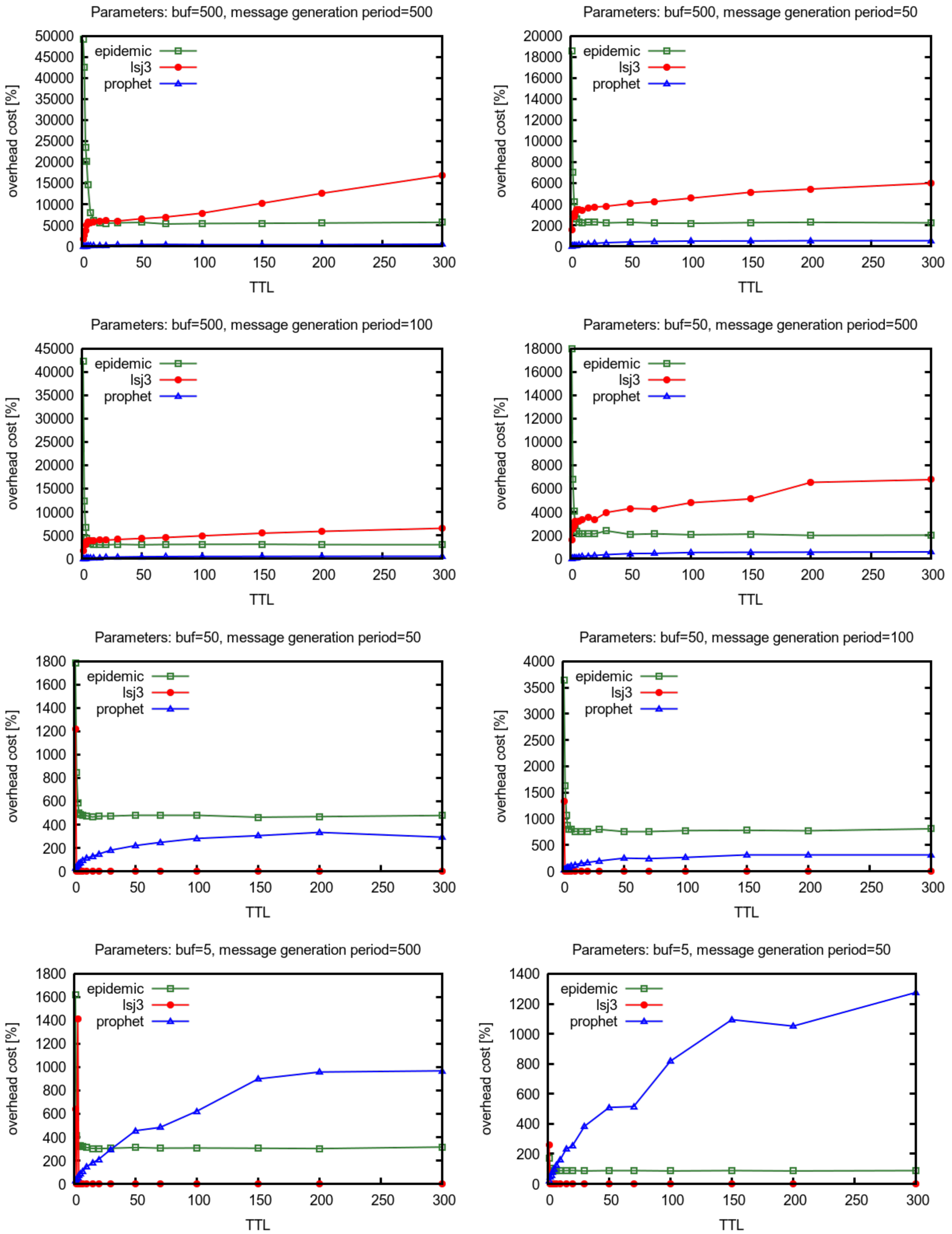

Figure 4.31: Simulation Scenario 3: Overhead cost ratio as a function of the the time to live

Main characteristics
1 region, urban area (Helsinki)
road density:high
Targets, nodes and node affiliations to targets are generated randomly at each step
No node mobility patterns

Figure 4.32: Simulation Scenario 1: The initial positions of nodes in OPN geographical area

Fig. 4.32 shows the initial positions of nodes in OPN geographical area.

4.4.4.1 Influence of Message Generation Period to Performance Metrics

Simulation Setup We conducted simulations for different of message generation period (referred as send period in graphs) and observed the influence of different message generation period to the performance metrics. The range of message generation periods was from 1 to 500. Lower values of message generation period imply higher rates of message generation by nodes, and consequently higher number of messages, which are simultaneously present in simulation. The other simulation parameters TTL and node message buffer were set as follows:

1. set of simulations: TTL = 500, buffer = 500

Parameter	Range
Map	Helsinki
Simulation size	2300 x 2300 m
Moving speed	random 0.5 - 1.5 m/s
Transmission range	20 m
Simulation Time	432000 time units ~ 10 days
Sampling Period T_s	0.5 time units ~ 1 second
Message size	36 bytes
Node Buffer Size	1 - 500 messages
Message Generation Period	10 - 500 time units ~ 20 - 1000 seconds
Time to live	1 - 300 transmissions

Table 4.4: Simulation Scenario 4 - Random: Simulation setup for ONE Simulator

2. set of simulations: TTL = 500, buffer = 50
3. set of simulations: TTL = 500, buffer = 50
4. set of simulations: TTL = 50, buffer = 500
5. set of simulations: TTL = 50, buffer = 100
6. set of simulations: TTL = 50, buffer = 50
7. set of simulations: TTL = 5, buffer = 500
8. set of simulations: TTL = 5, buffer = 50

Fig. 4.33 shows the message delivery ratio (referred as delivered messages [%] in graphs) as a function of the period of message generation. The range of observed message generation period was an interval from 1 to 500. Lower values of message generation period imply higher rates of message generation by nodes, and consequently higher number of messages, which are simultaneously present in simulation. As it is shown in graphs, the extremely low values of message generation period cause network congestion and the message delivery ratio rapidly decreases. Furthermore, it can be observed that the achieved results strongly depend on the size of message buffer. For the extremely short buffers (buffer = 5), the performance of the proposed method decreases and the message delivery ratio of HRC1 is about one percent. For the extremely short buffers (buffer = 5), The First Contact performance is the best. It corresponds to the random character of the scenario. The best results (about 80 percent) were achieved for the large buffer (buffer = 500). The results are similar for HRC1 and PROPHET for simulation scenarios, where buffer = 500. Both for the HRC1 routing scheme and PROPHET, the appropriate size of message buffer is crucial. These method do not perform well for extremely low values of message buffer.

For the First Contact routing, the value of TTL is crucial. We observe two types of dependencies between the message generation period and message delivery ratio. For the large values of the buffer parameter, the graph indicates rather logarithmic increase in message delivery ratio. For the small values of the buffer parameter, the graph indicates rather linear increase in message delivery ratio.

Fig. 4.34 shows the average message delivery delay in seconds (referred as the average delay in graphs) as the function of the message generation period. We observe two types of dependencies between the message generation period and the average message delivery delay. For the large values of the buffer, the graph indicates no influence of the message generation period to the average message delivery ratio. For the small values of the buffer parameter, the graph indicates high values of average message delivery delay for extremely small values of message generation period. As he period increases, the graph becomes flat. For the simulation scenarios with buffer=500, the average message delivery delay of the proposed method HRC2 is comparable to those ones of PROPHET and First Contact routing. The average message delivery delay of the epidemic routing is the best.

Fig. 4.35 shows the average number of hops as a function of the message generation period. In accordance to our assumptions, the high number of hopes can be observed when the First Contact routing scheme was in use in simulation. The average number of hopes of the proposed routing schema HRC2 is low and it is comparable to Epidemic routing. The graphs are almost flat. It can be interpreted as considering that the number of hopes depends particularly on the implemented routing method.

Fig. 4.36 shows the overhead cost ratio (referred as overhead cost in graphs) as the function of the message generation period. The graphs indicates that the proposed method has high overhead. The lover values of overhead of Epidemic routing are caused by the metrics which we use for overhead computation. Generated, but never set messages are not taken into account.

4.4.4.2 Influence of Buffer Size to Performance Metrics

We conducted simulations for simulation for different sizes of message buffer (from 1 to 500 messages) and observed the influence of message buffer size to the performance metrics. The other simulation parameters TTL and message generation period (referred as send period in graphs) were set as follows:

1. set of simulations: TTL = 500, message generation period = 500
2. set of simulations: TTL = 500, message generation period = 100
3. set of simulations: TTL = 500, message generation period = 50
4. set of simulations: TTL = 50, message generation period = 500
5. set of simulations: TTL = 50, message generation period = 100
6. set of simulations: TTL = 50, message generation period = 50

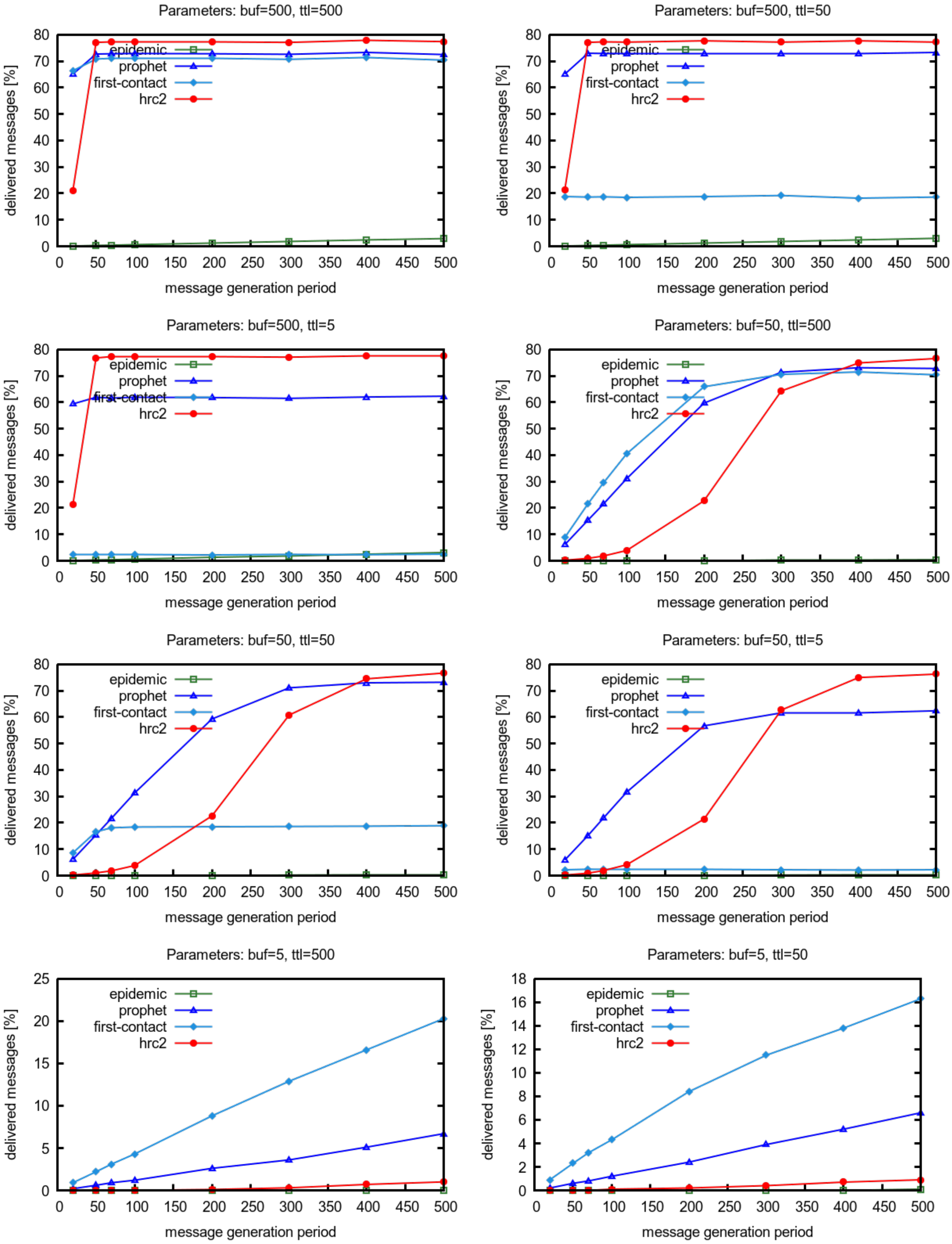

Figure 4.33: Simulation Scenario 4: Message delivery ratio as a function of the period of message generation

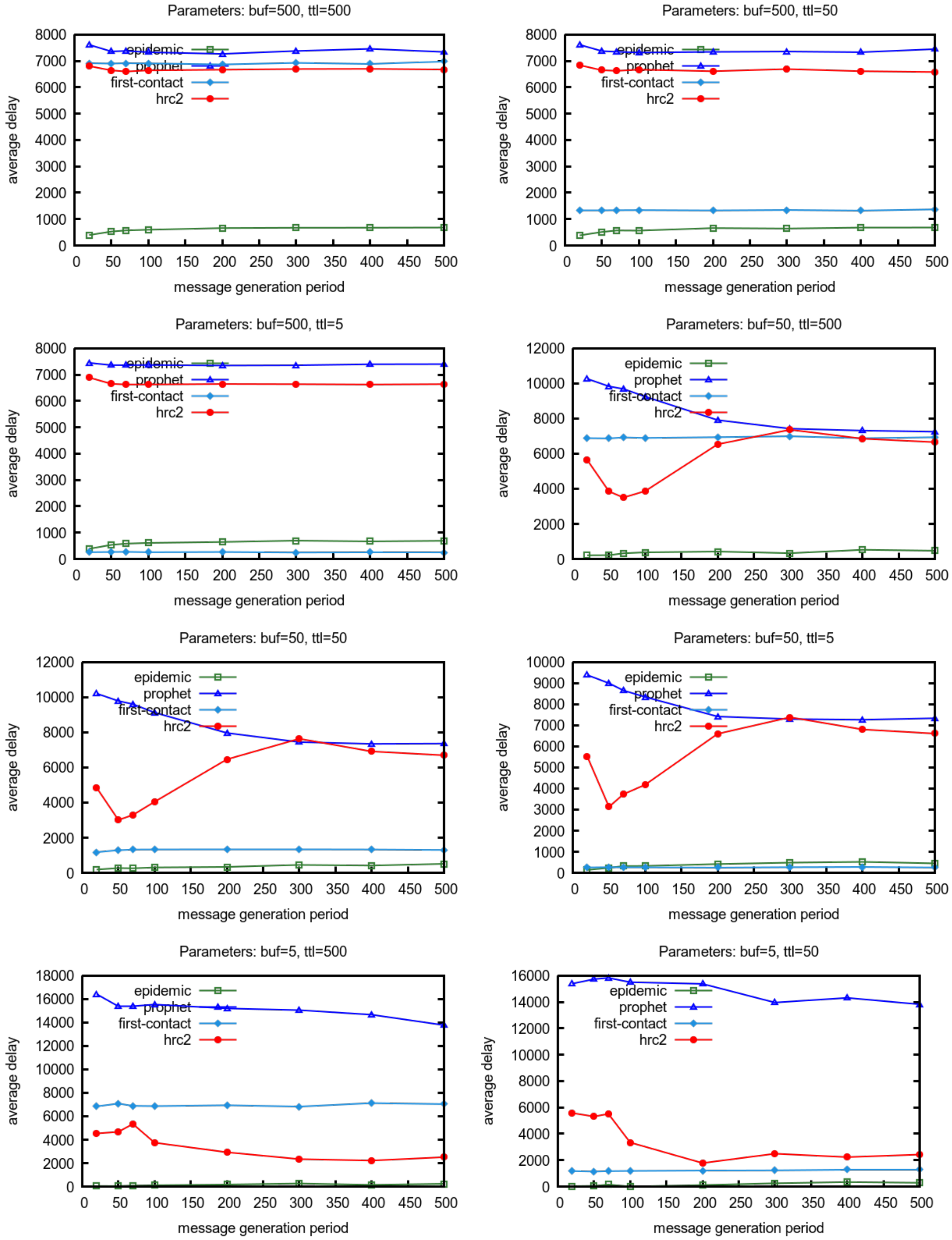

Figure 4.34: Simulation Scenario 4: Message delivery delay as a function of the period of message generation

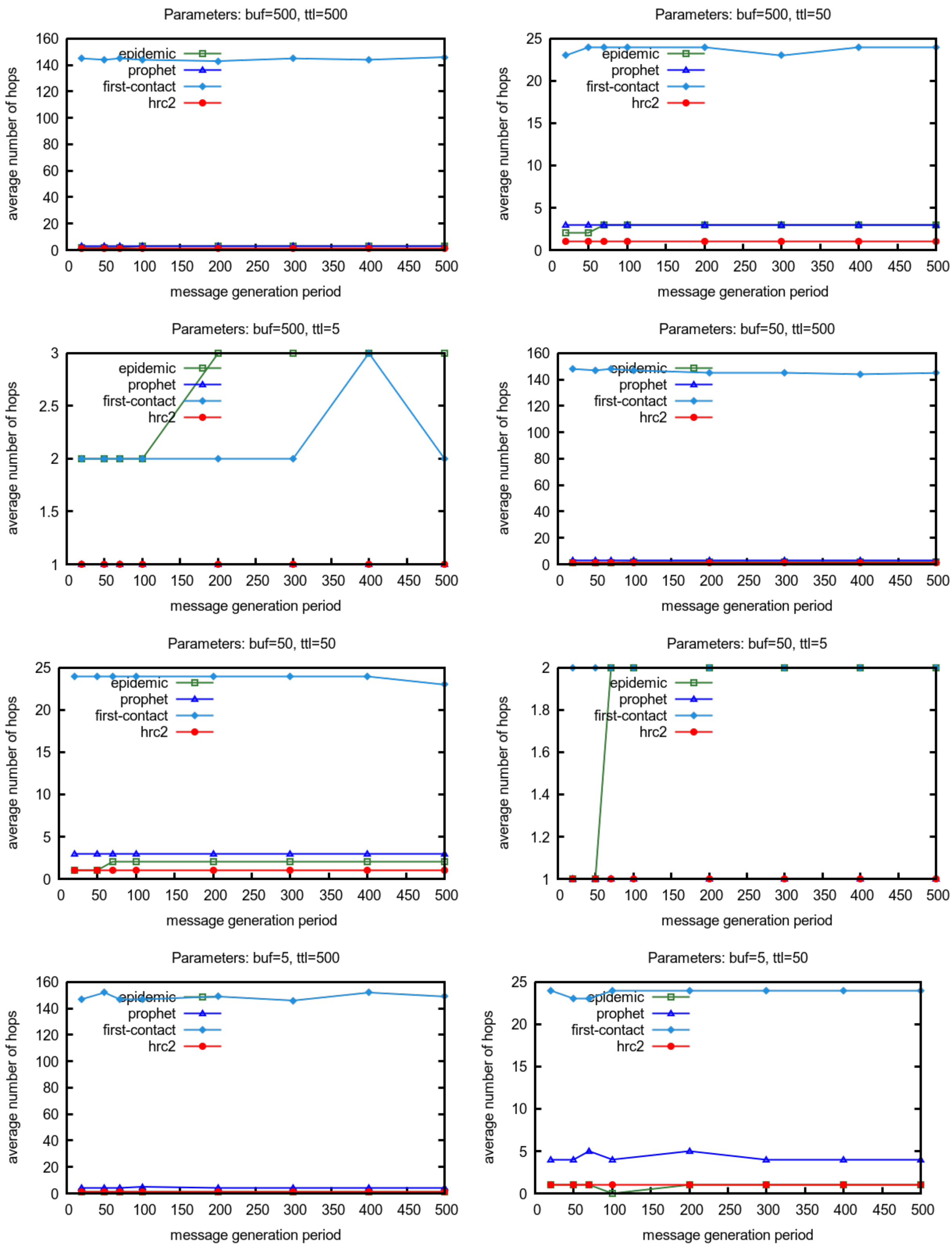

Figure 4.35: Simulation Scenario 4: Average number of hops as a function of the period of message generation

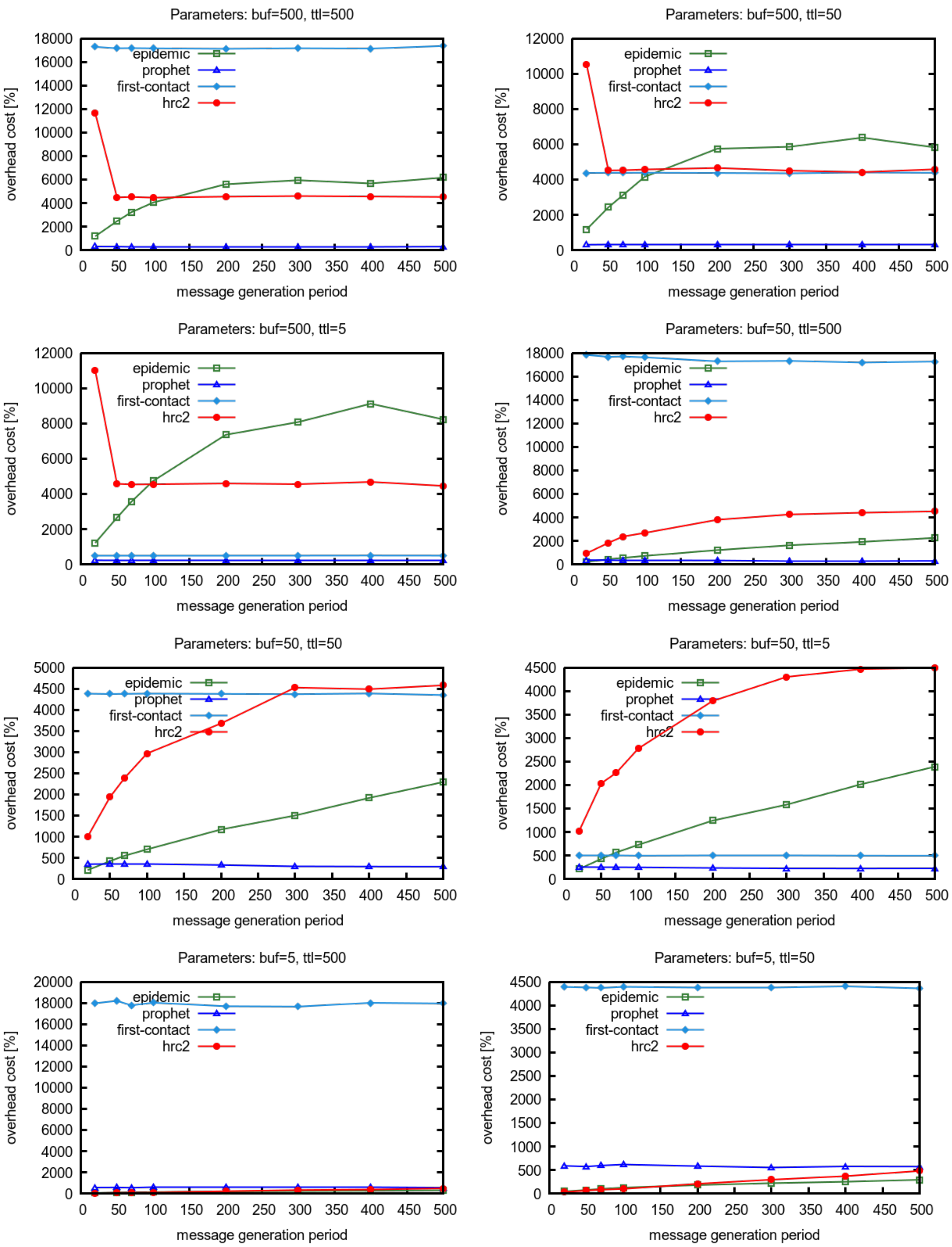

Figure 4.36: Simulation Scenario 4: Overhead cost ratio as a function of the period of message generation

7. set of simulations: TTL = 5, message generation period = 500

8. set of simulations: TTL = 5, message generation period = 50

Both the TTL and message generation periods are referred in time units of simulation. 1 time unit = 2 s.

Fig. 4.37 shows the message delivery ratio (referred as delivered messages [%] in graphs) as a function of the message buffer size. All the methods are comparable fr this random simulation scenario and the message delivery ratio reaches about 80 percent. For message generation period = 500 (1000s), the message delivery ratio of HRC2 and PROPHET grows as a function of the size of the message buffer until the buffer size is equal to 50 messages and than stays flat. For shorter message generation periods is the message delivery ratio dependent on the buffer size: the higher values of buffer size imply the higher values of message delivery ratio.

Fig. 4.38 shows the average message delivery delay (referred as average delay [%] in graphs) as a function of the message buffer size. The smaller values of buffer size have a strong influence on message delivery delay. Since the buffer size reaches a threshold value, no influence to message delivery delay is observable. The buffer threshold value depend on the values of parameters TTL and message generation period.

Fig. 4.39 shows the average number of hopes as a function of the message buffer size. The highest value of average number of hopes is achieved by the First Contact routing methods. No significant dependency on buffer size is observed.

Fig. 4.40 shows the overhead cost ratio (referred as overhead cost in graphs) as the function of the message buffer size. The graphs indicates that the proposed method preforms better than First Contact but worse than PROPHET. form the view of point of overhead, but it works without network congestion. The lover values of overhead of Epidemic routing are caused by the metrics which we use for overhead computation. Generated, but never set messages are not taken into account. We can observe unexpected extremely high peaks of overhead cost ratio for HRC2 routing scheme for several combinations of the values of TTL and message generation period and buffer size 200 or 400. The configuration with these parameters leads to computation of large communities. HRC2 routing scheme uses the approach, that the routing inside the communities is epidemic routing with timeout. The occurrences of peaks of overhead ratio can be eliminated by implementation of more tight constraints of application of epidemic routing into method or by limiting the number of nodes forming communities.

4.4.4.3 Influence of Time-to-live (TTL) to Performance Metrics

We conducted simulations for different values of TTL (from 1 to 300 time units). 1 simulation time unit is equal to 2 s. We observed the influence of TTL to the performance metrics. The other simulation parameters the size of node message buffer and message generation period (referred as send period in graphs) were set as follows:

1. set of simulations: buffer = 500 messages, message generation period = 500

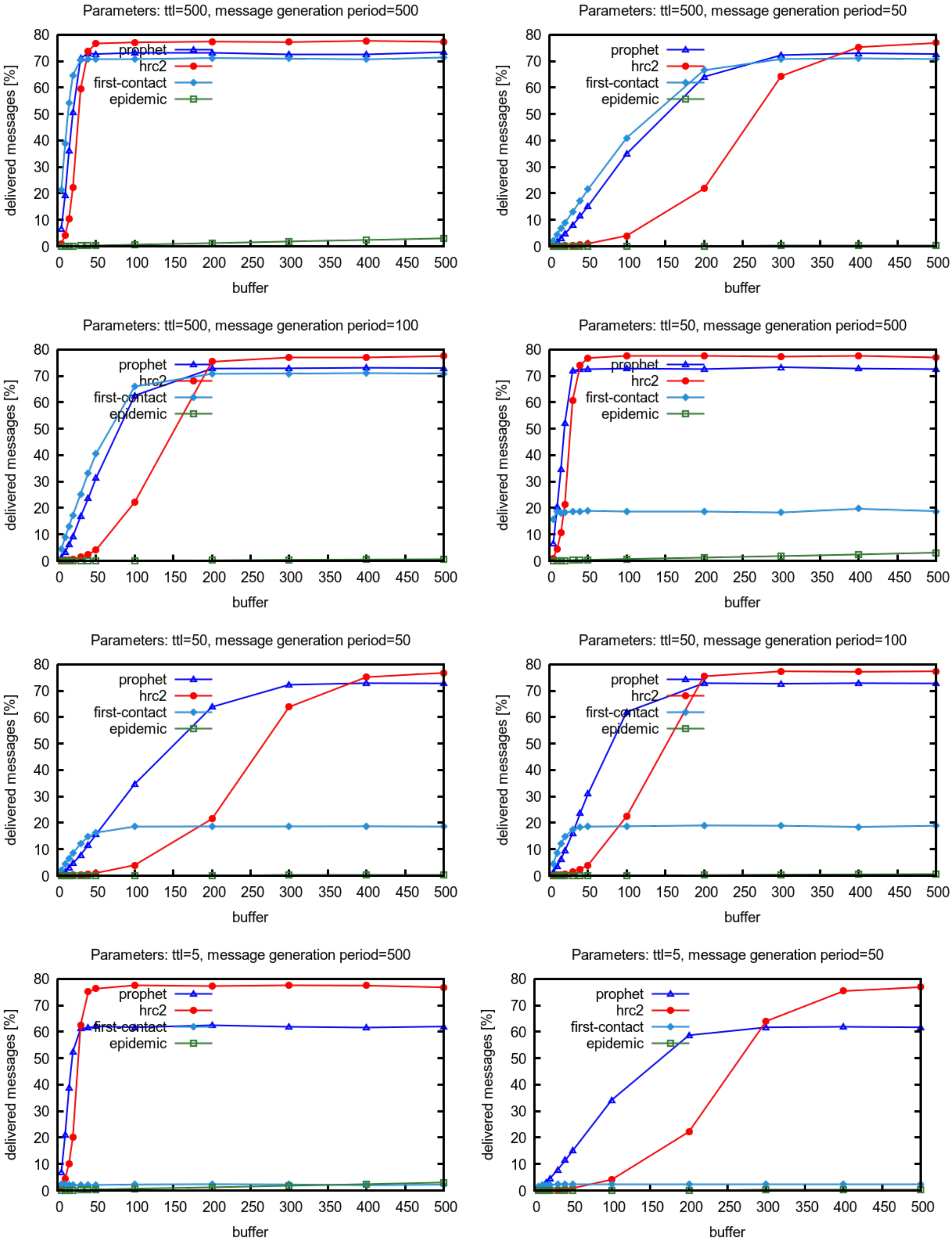

Figure 4.37: Simulation Scenario 4: Message delivery ratio as a function of the size of node message buffer

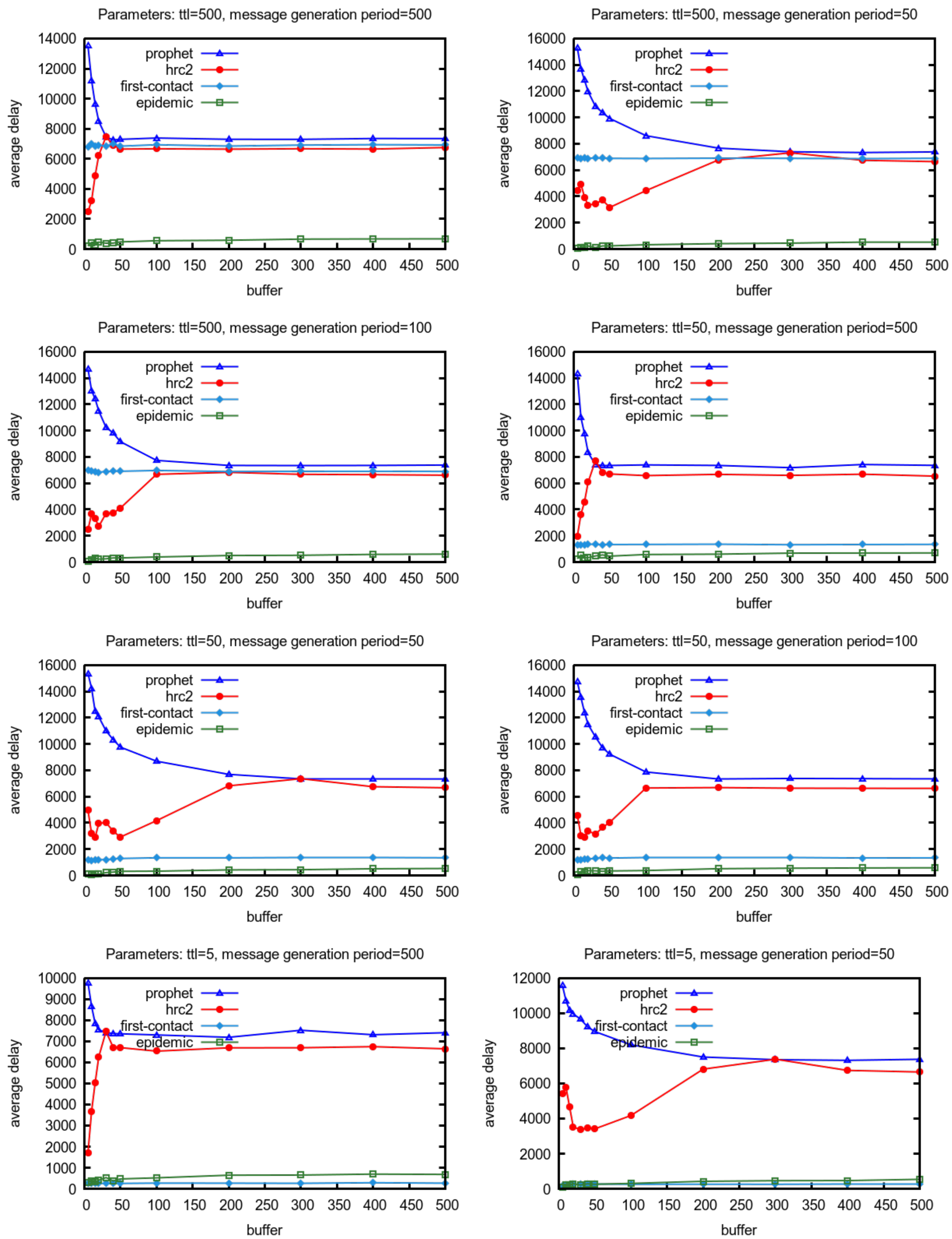

Figure 4.38: Simulation Scenario 4: Message delivery delay as a function of the size of node message buffer

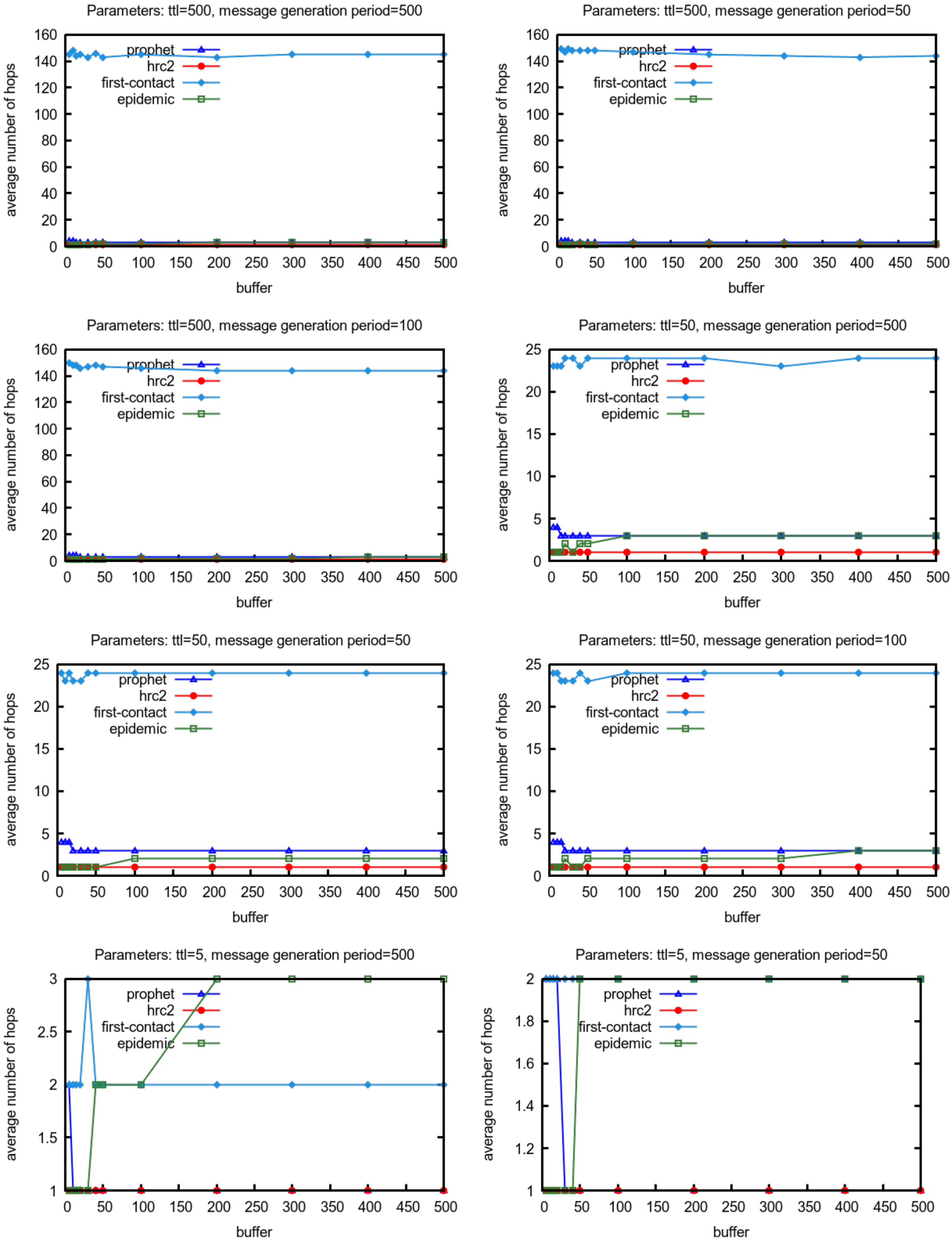

Figure 4.39: Simulation Scenario 4: Average number of hops as a function of the size of node message buffer

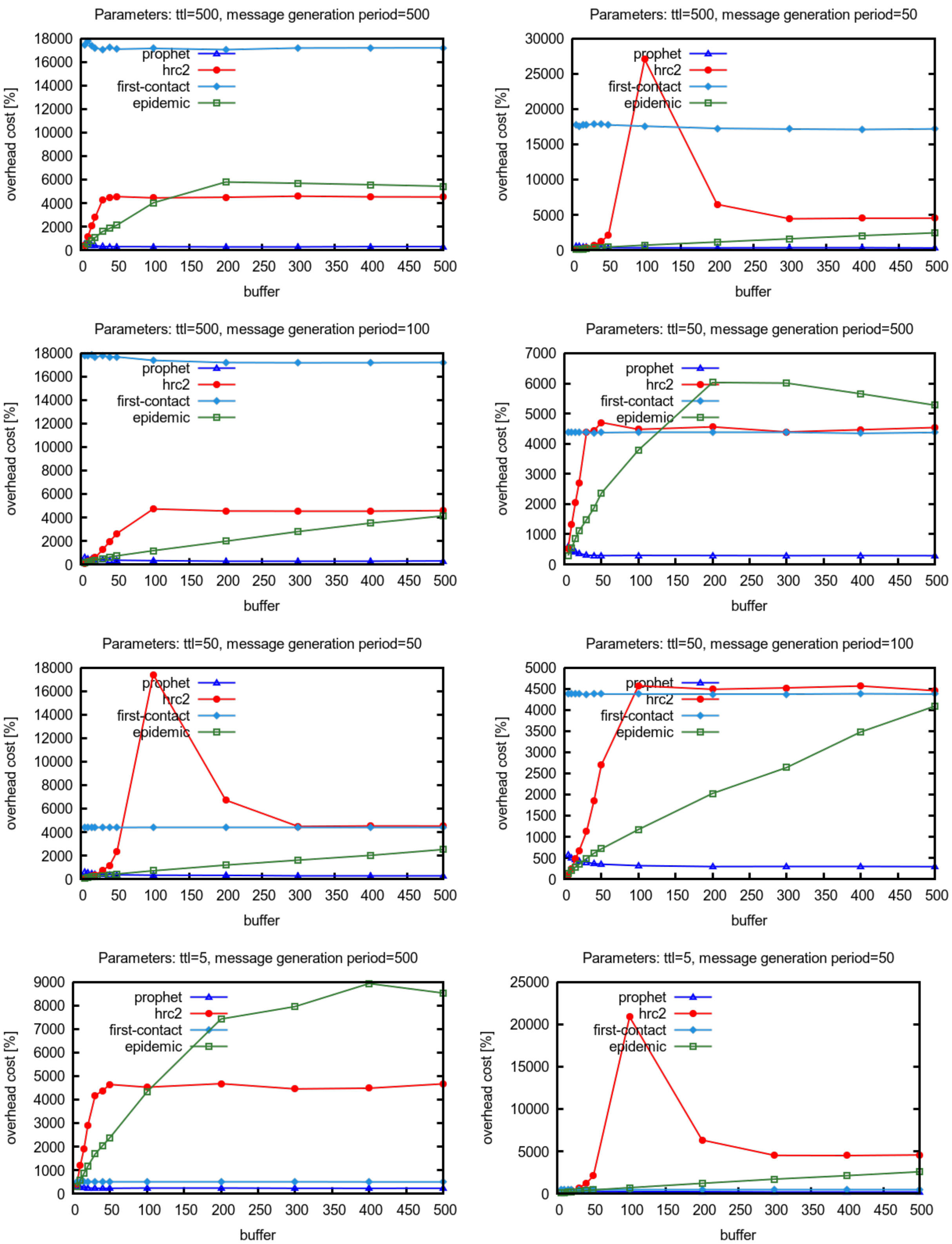

Figure 4.40: Simulation Scenario 4: Overhead cost ratio as a function of the size of node message buffer

2. set of simulations: buffer = 500 messages, message generation period = 100
3. set of simulations: buffer = 500 messages, message generation period = 50
4. set of simulations: buffer = 50 messages, message generation period = 500
5. set of simulations: buffer = 50 messages, message generation period = 100
6. set of simulations: buffer = 50 messages, message generation period = 50
7. set of simulations: buffer = 5 messages, message generation period = 500
8. set of simulations: buffer = 5 messages, message generation period = 50

Both the TTL and message generation periods are referred in time units of simulation. 1 time unit = 2 s.

Influence of TTL to message delivery ratio Fig. 4.41 shows the message delivery ratio (referred as delivered messages [%] in graphs) as function of TTL. The graphs indicates that HRC2 routing scheme proposed well and its performance in message delivery is comparable to PROPHET for higher values of message buffer. HRC2 doesn't not perform well for buffer size of 5. The proposed method performs well when the buffer is equal or higher than some threshold values.

Fig. 4.42 shows the average message delivery delay (referred as average delay in graphs) as the function of TTL. The graphs are flat. The graphs indicate low values of average message delivery delay for the small values of TTL, but the most analyzed methods have poor performance in message delivery for small values of TTL. It implies that these low message delivery delays for the small values of TTL are probably caused by the overall low number of delivered messages in the system.

Fig. 4.43 shows the average number of hops as function of TTL. In accordance to our assumptions, the graphs indicate linear dependence between the number of the hopes and TTL for the First Contact routing scheme, otherwise the graphs are almost flat. The average number of hopes of the proposed routing schema HRC2 is low and it is comparable to Epidemic routing and PROPHET. It can be interpreted as considering that the number of hopes doesn't not depend on TTL for the implemented routing method of HRC2, PROPHET or epidemic routing and that there is a linear dependency on TTL for the First Contact routing scheme.

Fig. 4.44 shows overhead cost ratio as a function of TTL. The graphs indicates no correlation between TTL and proposed method HRC2. For the First Contact the almost linear dependency of the overhead cost ratio to TTL is observed.

4.4.5 Experiment 5

Scenario: Simulation Scenario 5
Our tested method: ANMA
Compared to: Epidemic Routing, PROPHET, First Contact

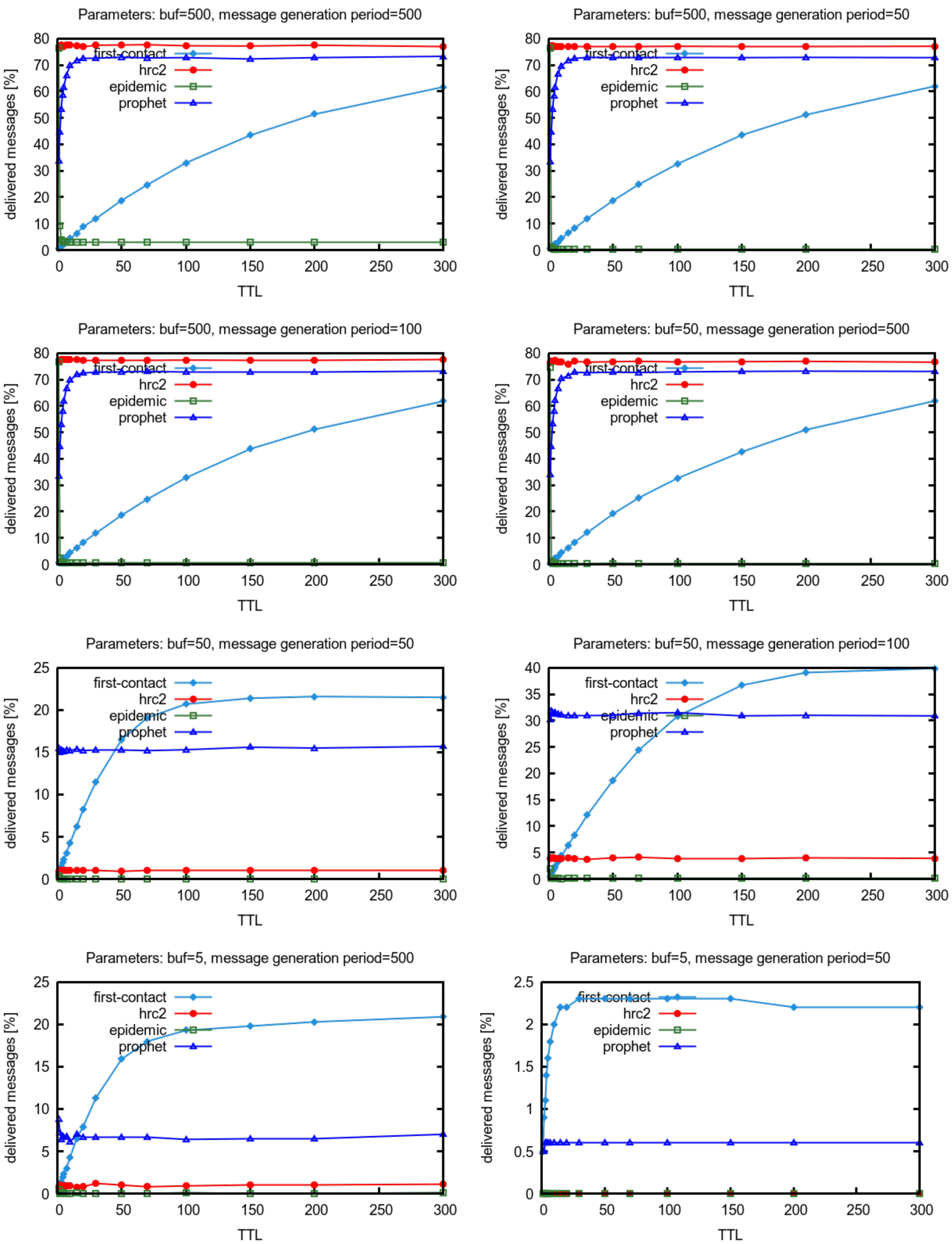

Figure 4.41: Simulation Scenario 4: Message delivery ratio as a function of the time to live

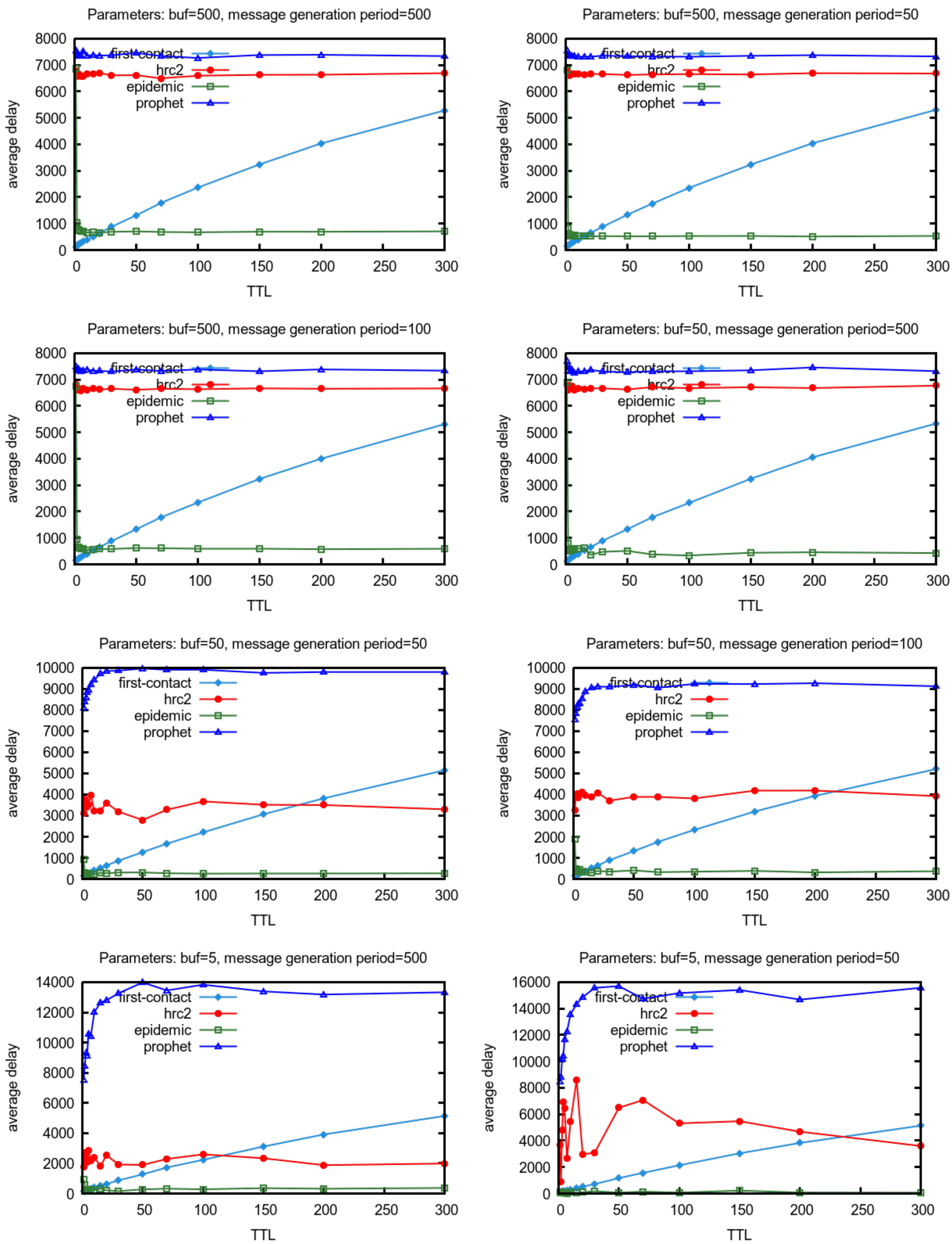

Figure 4.42: Simulation Scenario 4: Message delivery delay as a function of the time to live

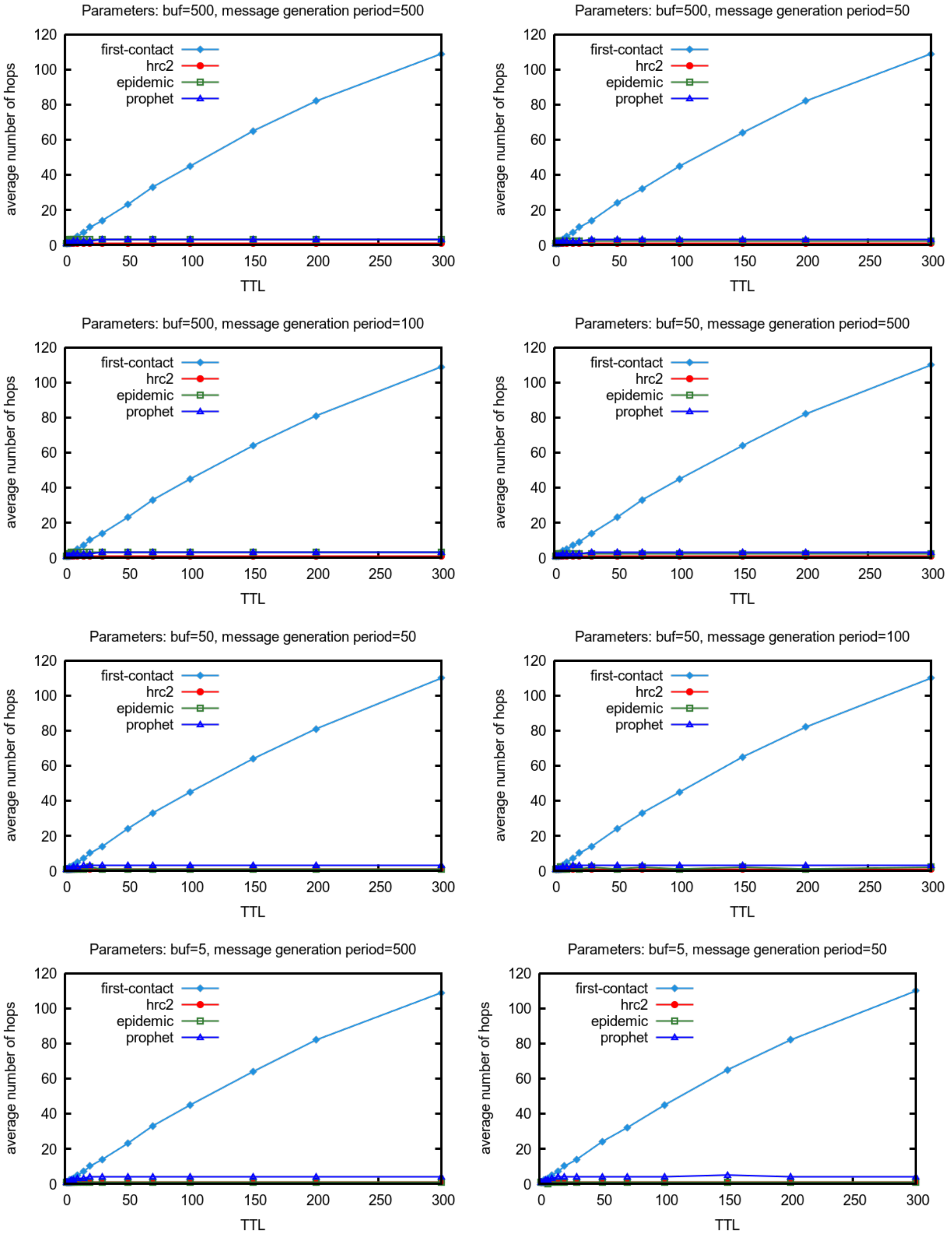

Figure 4.43: Simulation Scenario 4: Average number of hops as a function of the time to live

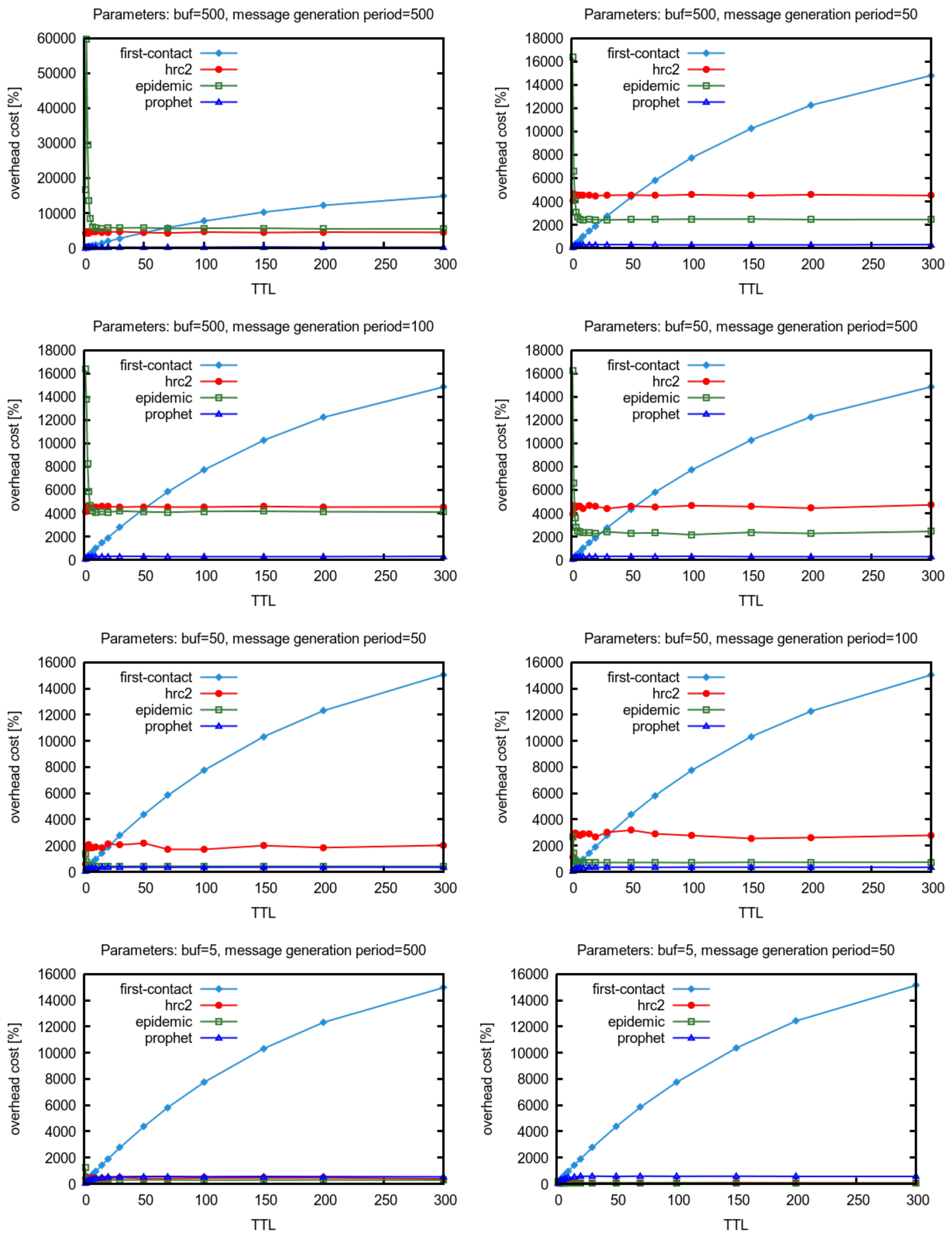

Figure 4.44: Simulation Scenario 4: Overhead cost ratio as a function of the the time to live

Four neighborhood geographical regions
Changing number of nodes visiting both regions
All nodes generate messages in predefined time period

Parameter	Range
Map	Venice
Simulation size	4500 x 3400 m
Moving speed	random 0.5 - 1.5 m/s
Transmission range	50 m
Simulation Time	432000 time units ~ 10 days
Sampling Period T_s	0.5 time units ~ 1 second
Message size	36 bytes
Node Buffer Size	50 messages
Message Generation Period	100 time units ~ 200 seconds
Time to live	5 transmissions

Table 4.5: Simulation Scenario 5: Simulation setup for ONE Simulator

This experiment was conducted on simulation scenario 5. The OPN consists of four square target neighborhood regions located in the urban area of Venice. The selected geographical area is characterized by the high density of roads and by the water channel which can be crossed by nodes only at some places (bridges). The node mobility pattern was designed in such a way, that there were two separated node groups at each time step. When the network changed behaviour, another two separated groups appear. We selected the uniform probabilistic distribution of node initial positions in each OPN geographical region and log-normal probabilistic distribution of node stays in targets during the day phase. The simulation was conducted for 6-day cycle with periodically changing day traffic pattern. For the purposes of OPN routing, we selected the data collected in time interval of 3 hours (7 AM to 10 AM) from each day of the 6-day simulation. The analyzed time interval was limited to 3 hours per day primary for the computational reasons. The selected 3-hour interval consists of mobility patterns changing each twenty minutes. The data from three days were used to train the model. The data from three days were used to test the performance of the proposed method. The number of nodes in simulation was 100 nodes. The simulation parameters were set as follows: message generation period=500 time units, buffer=200 messages, TTL=100 time units.

RESULTS: The following performance metrics were used: the message delivery ratio, average message delivery delay, average number of hopes and overhead cost. All compared methods achieve approximately the same message delivery delay. The average number of hopes of the proposed method for this simulation is low, about 3 hopes. The highest value of average hopes is achieved by First Contact routing protocol, where the average number of hopes was 19. The proposed method has a good overhead. The proposed method outperforms other tested methods in message delivery ratio, about 50 percent of delivered messages in comparison to 19 percent by Prophet and 10 percent of Epidemic Routing and First Contact. This reflects that the proposed ANMA routing scheme deliver messages even when the original OPN network schema contains two separated contact graphs at some time periods.

Chapter **5**

Conclusions

5.1 Summary

The dissertation thesis deals with the issue of special routing algorithms designed for communication in opportunistic networks. The opportunistic networks (OPN) are networks disseminating messages with the store- carry-forward routing principle. The key function of OPN routing protocols is to make decisions on message forwarding.

We proposed the routing metric combining utilization of geographical data and unsupervised machine learning (cluster analysis) called Hierarchical Routing with Clustering. Two versions of this routing algorithm were introduced: i) Hierarchical Routing with Clustering 1 (HRC1), ii) Hierarchical Routing with Clustering 2 (HRC2). These versions differ from each other in way how the knowledge about the detected geographic regions is represented. HRC1 uses the set representation of the geographic structure (the elements are detected geographic regions), while HRC2 uses the graph representation (the detected geographic regions are nodes, the edges represent the connections realized by the moving nodes of OPN). The novelty of this approach consists in combining the use of the knowledge about geographic structure of the network results of cluster analysis in routing algorithm with the knowledge obtained from the OPN contact graph. The routing metrics have been designed in order to select the most optimal nodes which have the highest probability to be a part of the paths of successfully delivered messages with respect to maximization of message delivery ratio and the minimization of message delivery delay. The performance of the proposed method was tested on several simulation scenarios and compared to four well-known routing protocols as Epidemic routing with the limited message buffer, PRoPHET, First Contact and BUBBLE-Rap. The simulations were conducted for different values of parameters TTL, Buffer Size, Message Generation Period and Number of Nodes. For the sufficient parameters of generated message period and buffer size and TTL, the proposed method achieves performance almost 90 percent of delivered messages and outperforms other implemented methods.

We proposed the routing metric which uses supervised machine learning technique (Support Vector Machine) as a part of decision making mechanism in OPNs with regular node mobility patterns. We proposed a SVM-based routing protocol. The main idea consists in dividing the area of the OPN into cells and train SVM classifier for each pair of the cells. During the training phase, the messages are spread over the network using epidemic routing. Each node has implemented a set of SVM classifiers. The simulation is the testing phase. We conducted experiments on simulation scenario, where the OPN consists of two separated target regions located in the urban area of Venice. The selected geographical area is characterized by the high density of roads and by the water channel which can be crossed by nodes only at some places (bridges). The method performance was about 40 percent in message delivery and outperforms other methods.

We proposed the routing method, which continuously evaluates the network state and can enhance the routing process by active changes in node behaviour. This routing scheme combines GMRF (Gaussian Random Fields) and ANMA (Active Node Movement Algorithm). The proposed method was tested on simulation scenario 5. The OPN consists of four square target neighborhood regions located in the urban area of Venice. The selected geographical area is characterized by the high density of roads and by the water channel which can be crossed by nodes only at some places (bridges). The node mobility pattern was designed in such a way, that there were two separated node groups at each time step. When the network changed behaviour, another two separated groups appear. We selected the uniform probabilistic distribution of node initial positions in each OPN geographical region and log-normal probabilistic distribution of node stays in targets during the day phase. The simulation was conducted for 6-day cycle with periodically changing day traffic pattern. In our implementation, we used centrally computed network evaluation. The proposed method outperforms other methods and in message delivery ratio, about 50 percent of delivered messages in comparison to 19 percent by Prophet and 10 percent of Epidemic Routing and First Contact.

The goals of dissertation thesis have been achieved.

5.2 Contributions of the Dissertation Thesis

The main contributions consist in:

1. We proposed the routing protocol Hierarchical Routing with Clustering, which combines three strategies in order to improve routing in OPNs. At the highest level, the proposed routing schema uses the knowledge about the geographical structure of the OPN. This geographical structure of the OPN is primary defined by the node targets (places visited by nodes) and by the node trajectories. This knowledge is extracted from the data by application of cluster analysis to triangulated an pre-processed data. At the middle level, the routing schema uses the knowledge of communication community constructed from the contact graph of the OPN. At the lowest level, e.g. inside the community the flooding routing schema is used for message dissemination. We proposed two versions of HRC. These versions differ from each other in way how the knowledge about the detected geographic regions is represented. HRC1 uses the set representation of the

geographic structure (the elements are detected geographic regions), while HRC2 uses the graph representation (the detected geographic regions are nodes, the edges represent the connections realized by the moving nodes of OPN). The novelty of this approach consists in combining the use of the knowledge about geographic structure of the network results of cluster analysis in routing algorithm with the knowledge obtained from the OPN contact graph.

2. We proposed how use the Support Vector Machines to make decisions about routing in OPNs with regular node mobility patterns. We proposed a SVM-based routing protocol. The main idea consists in dividing the area of the OPN into cells and train SVM classifier for each pair of the cells. During the training phase, the messages are spread over the network using epidemic routing. When the node encounters another node and intents to copy a message, it collects information on cell, where the nodes encountered, encountered node, destination node ID and time slot, when nodes encountered and the message ID. The destination node collects data containing message ID, sector, where the destination node is located, and the ID of the last hop node. The training data are collected for each sector. Each node has implemented a set of SVM classifiers.

3. We proposed and tested an active node behavior algorithm and verify that the decision making about a place of message passing can influence the delivery of messages to the destination node. We designed the Active node movement algorithm (ANMA). The experimental results were compared to the results obtained by the existing standard routing methods: Epidemic routing with the limited message buffer and Prophet routing. The novelty of this approach consists in combination of sector-based models of network behavior using GMRF and the feature of active decisions made by nodes about the node deviation from its planned route.

5.3 Future Work

One of the directions of the future research is closer connection of enhanced routing algorithms with GIS information systems and real world traffic data databases. The simulation environment, which we used, allows importing data about topography and infrastructure from public GIS databases. In the future, the incorporation of precious geographic data from public GIS systems as Open Topography GIS or OpenStreetMap could bring benefits. Open Topography GIS is a database, which contains precious topographical data. OpenStreetMap is a database which contains data for the whole world. These data includes points of interest, buildings, roads and road names, ferry routes. Unfortunately, in general, current public GIS databases do not yet offer traffic information. Although some cities publish GIS traffic data, this is still the activity of local authorities, not a general concept. In general, publishing GIS data on transport and traffic is not very widespread. Many data traffic databases are often private. Using enhanced routing algorithms on big data could bring a number of new problems. One of the problems to solve when working on big data would be, for example, a problem on selecting information stored in nodes.

A node storage capacity is limited. When working on big data, the nodes would not be even able remember identifiers of all other nodes in a opportunistic network. Also, many problems concerning a central repository and a central system for the collection and integration of big-data on traffic stay opened. In our work, we assumed the implicit cooperative behavior of nodes. Another interesting approach would be application on multi-agent systems, where agents set up contracts for message delivery. Multi-agent system is a software system, which consists of multiple interacting intelligent agents within an environment. Future research could focus on concluding contracts between opportunist nodes and an application of game theory.

Bibliography

[1] Abdulla, M. and Simon, R. *Analysis of Core-Assisted Routing in Opportunistic Networks.* Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, MASCOTS '07, 15th International Symposium, pp. 387–394, 2007.

[2] Ahmad, Khaleel, Udzir, Nur Izura, and Deka, Ganesh Chandra. Opportunistic Networks: Mobility Models, Protocols, Security, and Privacy. CRC Press, 2018.

[3] Alexander, Lauren, Jiang, Shan, Murga, Mikel, and González, Marta C. *Origin–destination trips by purpose and time of day inferred from mobile phone data.* Transportation research part c: emerging technologies, 58:240–250, 2015.

[4] Ariyakhajorn, J. et al. *A Comparative Study of Random Waypoint and Gauss-Markov Mobility Models in the Performance Evaluation of MANET.* Communications and Information Technologies, 2006. ISCIT '06. International Symposium, pp. 894–899, 2006.

[5] Ashbrook, Daniel and Starner, Thad. *Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users.* Personal Ubiquitous Comput., 7(5):275–286, 2003. ISSN 1617-4909. doi:10.1007/s00779-003-0240-0.

[6] Balasubramanian, Aruna, Levine, Brian, and Venkataramani, Arun. *DTN routing as a resource allocation problem.* ACM SIGCOMM Computer Communication Review, 37(4):373–384, 2007.

[7] Banerjee, S. and Misra, A. *Minimum energy paths for reliable communication in multi-hop wireless networks.* Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing, Lausanne, Switzerland, pp. 146–156, 2002.

[8] Basagni, Stefano, Chlamtac, Imrich, Syrotiuk, Violet R., and Woodward, Barry A. *A Distance Routing Effect Algorithm for Mobility (DREAM).* In Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Com-

puting and Networking, MobiCom '98, pp. 76–84. New York, NY, USA: ACM, 1998. ISBN 1-58113-035-X. doi:10.1145/288235.288254. Available from: `http://doi.acm.org/10.1145/288235.288254`

[9] Bavelas, Alex. *Communication Patterns in Task-Oriented Groups.* Journal of the Acoustical Society of America, 22:725–730, 1950. doi:10.1121/1.1906679.

[10] Bazzani, Armando, Giorgini, Bruno, Rambaldi, Sandro, Gallotti, Riccardo, and Giovannini, Luca. *Statistical laws in urban mobility from microscopic GPS data in the area of Florence.* Journal of Statistical Mechanics: Theory and Experiment, 2010(05):P05001, 2010.

[11] Becker, R. A., Caceres, R., Hanson, K., Loh, J. M., Urbanek, S., Varshavsky, A., and Volinsky, C. *A Tale of One City: Using Cellular Network Data for Urban Planning.* IEEE Pervasive Computing, 10(4):18–26, 2011. ISSN 1536-1268. doi:10.1109/MPRV.2011.44.

[12] Bettstetter, C. *Smooth is Better than Sharp: A Random Mobility Model for Simulation of Wireless Networks.* Proc. ACM International Workshop on Modeling Analysis, and Simulation of Wireless and Mobile Systems (MSWiM), 2001.

[13] Bigwood, Greg and Henderson, Tristan. *Bootstrapping opportunistic networks using social roles*, 2011. doi:10.1109/WoWMoM.2011.5986139. [online]. 2011. [Accessed 2019-03-29]. Available from: `https://www.researchgate.net/publication/221567104_Bootstrapping_opportunistic_networks_using_social_roles`

[14] Bigwood, Greg, Henderson, Tristan, and Bhatti, Saleem. *Social Delay-Tolerant Network Routing*, 2011. [online]. 2011. [Accessed 2019-03-29]. Available from: `https://www.researchgate.net/publication/228427704_Social_Delay-Tolerant_Network_Routing`

[15] Bigwood, Greg, Henderson, Tristan, Rehunathan, Devan, Bateman, Martin, and Bhatti, Saleem. *CRAWDAD dataset st_andrews/sassy (v. 2011-06-03).* Downloaded from `https://crawdad.org/st_andrews/sassy/20110603`, 2011. doi:10.15783/C7S59X.

[16] Blazevic, L. et al. *Self organized terminode routing simulation.* Proceedings of the 4th ACM international workshop on Modeling, analysis and simulation of wireless and mobile systems, pp. 81–88, 2001.

[17] Boldrini, C. et al. *Hi-BOp: a History-based Routing Protocol for Opportunistic Networks.* Proc. of IEEE WoWMoM 2007, 2007.

[18] Boldrini, C. et al. *HiBOp: a History Based Routing Protocol for Opportunistic Networks, Applications and Worksharing.* CollaborateCom 2007 - International Conference, pp. 150–159, 2007.

[19] Boldrini, C. et al. *Context and resource awareness in opportunistic network data dissemination.* World of Wireless, Mobile and Multimedia Networks, WoWMoM 2008, International Symposium, pp. 1–6, 2008.

[20] Boldrini, C. et al. *HiBOp: a History Based Routing Protocol for Opportunistic Networks.* Ubiquitous and Future Networks, ICUFN 2009, 2009.

[21] Bononi, L. et al. *Parallel and Distributed Simulation of Wireless Vehicular Ad Hoc Networks.* Proc. ACM/IEEE International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM), 2006.

[22] Borrego, Carlos, Sánchez-Carmona, Adrián, Li, Zhiyuan, and Robles, Sergi. *Explore and wait: A composite routing-delivery scheme for relative profile-casting in opportunistic networks.* Computer Networks, 123:51–63, 2017.

[23] Bradley, M., Bowman, J. L., and Griesenbeck, B. *Development and application of the SACSIM activity-based model system.* In 11th World Conference on Transport ResearchWorld Conference on Transport Research Society., 2007.

[24] Broch, J. et al. *A performance comparison of multi-hop wireless ad hoc network routing protocols.* Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (Mobicom98), 1998.

[25] Burleigh, S. et al. *Delay-Tolerant Networking - An Approach to Interplanetary Internet.* IEEE Communications Magazine, 41(6):128–136, 2003.

[26] Burns, B., Brock, O., and Levine, B. N. *MV Routing and Capacity Building in Disruption Tolerant Networks.* Proc. IEEE INFOCOM 2005, Miami, FL, 2005.

[27] Cadger, F., Curran, K., Santos, J., and Moffett, S. *A Survey of Geographical Routing in Wireless Ad-Hoc Networks.* IEEE Communications Surveys Tutorials, 15(2):621–653, 2013. ISSN 1553-877X. doi:10.1109/SURV.2012.062612.00109.

[28] Cahkeres, I.D. and Klein-Berndt, L. *AODVjr, AODV Simplified.* ACM SIGMOBILE Mobile Computing and Communications Review, 6(3):100–101, 2002. [online]. 2002. [Accessed 2018-01-22]. Available from: `http://moment.cs.ucsb.edu/pub/aodvjr.pdf`

[29] Caleffi, M. and Paura, L. *Opportunistic Routing for Disruption Tolerant Networks.* Advanced Information Networking and Applications Workshops, WAINA '09, International Conference, pp. 826–831, 2009.

[30] Cambridge University Engineering Department. *The Hidden Markov Model Toolkit (HTK)*, 2016. [online]. 2016. [Accessed 2018-01-22]. Available from: `http://htk.eng.cam.ac.uk/`

[31] Castiglione, Joe, Bradley, Mark, and Gliebe, John. Activity-Based Travel Demand Models: A Primer. Washington, DC: The National Academies Press, 2014. doi: 10.17226/22357. Available from: https://www.nap.edu/catalog/22357/activity-based-travel-demand-models-a-primer

[32] Cerf, V. and Kahn, R. *A Protocol for Packet Network Intercommunication.* IEEE Trans. on Comm., 22(5), 1974.

[33] Cerf, V. et al. *Delay-Tolerant Networking Architecture.* INTERNET-DRAFT, DTN Research Group, 2006. [online]. 2006. [Accessed 2018-01-22]. Available from: http://tools.ietf.org/pdf/draft-irtf-dtnrg-arch-08.pdf

[34] Chaintreau, A. et al. *Impact of Human Mobility on the Design of Opportunistic Forwarding Algorithms.* Proc. 26th Annual IEEE Conference on Computer Communications (INFOCOM'06), 2006.

[35] Chellappa, R. and Chatterjee, S. *Classification of textures using Gaussian Markov random fields.* IEEE Transactions on Acoustics, Speech and Signal Processing, 33(4):959–963, 1985.

[36] Chen, B. et al. *Span: an energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks.* Wireless Networks, 8(5):481–494, 2002.

[37] Chen, L.-J. et al. *A content-centric framework for effective data dissemination in opportunistic networks.* Selected Areas in Communications, IEEE Journal, 26(5):761–772, 2008.

[38] Chen, W., Letaief, K.B., and Cao, Z. *Opportunistic Network Coding for Wireless Networks Communications.* ICC'07 - IEEE International Conference, pp. 4634–4639, 2007.

[39] Chen, Yang, Borrel, Vincent, Ammar, Mostafa, and Zegura, Ellen. *A framework for characterizing the wireless and mobile network continuum.* ACM SIGCOMM Computer Communication Review, 41(1):5–13, 2011.

[40] Cheng, Tao and Anbaroglu, Berk. *Spatio-Temporal Clustering of Road Network Data.* In Fu Lee Wang, Hepu Deng, Yang Gao, and Jingsheng Lei, editors, Artificial Intelligence and Computational Intelligence, pp. 116–123. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. ISBN 978-3-642-16530-6.

[41] Choi, N. et al. *A Unified Framework of Node Mobility Models.* Military Communications Conference, 2007. MILCOM 2007, IEEE, pp. 1–7, 2007.

[42] Cichocki, A. et al. Adaptive Blind Signal and Image Processing. Wiley; 1 edition, 2002. ISBN 978-0471607915.

[43] CIRIELLO, G. *A review on models and algorithms for motif discovery in protein-protein interaction networks.* Brief. in Func. Genomics and Proteomic, 7:147–156, 2008.

[44] CLAUSET, Aaron, NEWMAN, Mark EJ, and MOORE, Cristopher. *Finding community structure in very large networks.* Physical review E, 70(6):066111, 2004.

[45] CRISTIANINI, N. and SHAWE-TAYLOR, J. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press, 2000. ISBN 0-521-78019-5.

[46] DANGALCHEV, Chavdar. *Residual closeness in networks.* Physica A: Statistical Mechanics and its Applications, 365(2):556 – 564, 2006. ISSN 0378-4371. doi:https://doi.org/10.1016/j.physa.2005.12.020.

[47] DANON, Leon, DÍAZ-GUILERA, Albert, DUCH, Jordi, and ARENAS, Alex. *Comparing community structure identification.* Journal of Statistical Mechanics: Theory and Experiment, 2005(09):P09008–P09008, 2005. doi:10.1088/1742-5468/2005/09/p09008.

[48] DEGUCHI, K. *Two-dimensional auto-regressive model for analysis and sythesis of gray-level textures.* Proceedings of the First International Symposium for Science on Form, 1986. [online]. 1986. [Accessed 2018-01-22]. Available from: `http://www.scipress.org/e-library/sof/pdf/0441.PDF`

[49] DEKKER, Anthony. *Conceptual Distance in Social Network Analysis.* Journal of Social Structure (JoSS), 6, 2005.

[50] FAN, K. et al. *A Network-Coding Method to Improve the Performance of Opportunistic Networks.* WiCom '09. 5th International Conference on Wireless Communications, Networking and Mobile Computing, pp. 1–3, 2009.

[51] FAY, Damien, KUNEGIS, Jérôme, and YONEKI, Eiko. *Centrality and Mode Detection in Dynamic Contact Graphs; a Joint Diagonalisation Approach.* In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '13, pp. 41–48. New York, NY, USA: ACM, 2013. ISBN 978-1-4503-2240-9. doi:10.1145/2492517.2492540. Available from: `http://doi.acm.org/10.1145/2492517.2492540`

[52] FAY, Damien, KUNEGIS, Jérôme, and YONEKI, Eiko. *Centrality and mode detection in dynamic contact graphs; a joint diagonalisation approach.* In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 41–48. ACM, 2013.

[53] FIORE, M., OLARIU, S., and WEIGLE, M.C. Vehicular Mobility and Network Simulation. Taylor and Francis, 2008.

[54] Fiore, M. et al. *Understanding Vehicular Mobility in Network Simulation.* Proc. 1st IEEE Workshop on Mobile Vehicular Networks (MoveNet'07, in conjunction with IEEE MASS'07), 2007.

[55] Fournet, Julie and Barrat, Alain. *Contact patterns among high school students.* PloS one, 9(9):e107878, 2014.

[56] Freeman, Linton. *A Set of Measures of Centrality Based on Betweenness.* Sociometry, 40:35–41, 1977. doi:10.2307/3033543.

[57] Gamal, A. E. and Mammen, J. *Throughput-delay trade-off in wireless networks.* Proc. IEEE INFOCOM, pp. 464–475, 2004.

[58] Gao, Wei, Li, Qinghua, Zhao, Bo, and Cao, Guohong. *Multicasting in delay tolerant networks: a social network perspective.* In Proceedings of the tenth ACM international symposium on Mobile ad hoc networking and computing, pp. 299–308. ACM, 2009.

[59] Génois, Mathieu and Barrat, Alain. *Can co-location be used as a proxy for face-to-face contacts?* EPJ Data Science, 7(1):11, 2018.

[60] Goodman, D. et al. *INFOSTATIONS: A New System Model for Data and Messaging Services.* IEEE VTC'97, 2:969–973, 1997.

[61] Greede, A., Allen, S.M., and Whitaker, R.M. *RFP: Repository Based Forwarding Protocol for Opportunistic Networks.* Next Generation Mobile Applications, Services and Technologies, NGMAST '09. Third International Conference, pp. 329–334, 2009.

[62] Grossglauser, M and Tse, D. N. C. *Mobility Increases the Capacity of Ad-Hoc Wireless Networks.* IEEE/ACM Trans. Net., 10(4):477–486, 2002.

[63] Gu, Bo and Hong, Xiaoyan. *Capacity-aware routing using throw-boxes.* In 2011 IEEE Global Telecommunications Conference-GLOBECOM 2011, pp. 1–5. IEEE, 2011.

[64] Guan, Jianfeng, Chu, Qi, and You, Ilsun. *The social relationship based adaptive multi-spray-and-wait routing algorithm for disruption tolerant network.* Mobile Information Systems, 2017, 2017.

[65] Gupta, P. and Kumar, P. R. *The Capacity of Wireless Networks.* IEEE Trans. Info. Theory, 46(2):388–404, 2000.

[66] Harri, J. et al. *On Meaningful Parameters for Routing in VANETs Urban Environments under Realistic Mobility Patterns.* Proc. 1st IEEE Workshop on Automotive Networking and Applications (AutoNet'06, in conjunction with IEEE Globecom 2006), 2006.

[67] HARRI, Jerome, FILALI, Fethi, and BONNET, Christian. *Mobility models for vehicular ad hoc networks: a survey and taxonomy.* IEEE Communications Surveys & Tutorials, 11(4), 2009.

[68] HELBING, D. *Traffic and Related Self-driven Many-particles Systems.* Rev. Modern Physics, 73:1067–1141, 2001.

[69] HELMY, A. *MobiLib: Community-wide Library of Mobility and Wireless Networks Measurements*, 2017. [online]. 2017. [Accessed 2018-01-22]. Available from: `https://www.cise.ufl.edu/~helmy/MobiLib.htm`

[70] HOLME, P. and SARAMAKI, J. Temporal Networks: The Cabspotting Project. Springer, 2013. ISBN 978-3-642-36461-7.

[71] HOQUE, Mohammad Asadul, HONG, Xiaoyan, and DIXON, Brandon. *Analysis of mobility patterns for urban taxi cabs.* In 2012 international conference on computing, networking and communications (ICNC), pp. 756–760. IEEE, 2012.

[72] HSU, Ch., CHANG, Ch., and LIN, Ch. *A Practical Guide to Support Vector Classification.* Technical report, Department of Computer Science and Information Engineering, National Taiwan University, 2003.

[73] HSU, W. et al. *Weighted Waypoint Mobility Model and its Impact on Ad Hoc Networks.* Proc. ACM Mobile Computer Communications Review (MC2R), 2005.

[74] HUANG, C.-M., LAN, K.-C., and TSAI, C.-Z. *A Survey of Opportunistic Networks.* Advanced Information Networking and Applications - Workshops, AINAW 2008, 22nd International Conference, pp. 1672–1677, 2008.

[75] HUANG, J.-H. et al. *Improving Opportunistic Data Dissemination via Known Vector.* Advanced Information Networking and Applications Workshops, WAINA '09, International Conference, pp. 838–843, 2009.

[76] HUI, Pan, CROWCROFT, Jon, and YONEKI, Eiko. *Bubble rap: Social-based forwarding in delay-tolerant networks.* IEEE Transactions on Mobile Computing, 10(11):1576–1589, 2011.

[77] HYUNUK, K. and SONG, H.Y. *Formulating Human Mobility Model in a Form of Continuous Time Markov Chain.* Procedia CS, 10, 2012.

[78] INTERNET RESEARCH TASK FORCE (IRTF). *RFC 6693*, 2012. [online]. 2012. [Accessed 2018-01-22]. Available from: `https://tools.ietf.org/html/rfc6693`

[79] INTERPLANETARY NETWORKING SPECIAL INTEREST GROUP. *InterPlanetary Internet Project*, 2018. [online]. 2018. [Accessed 2018-01-22]. Available from: `http://www.ipnsig.org`

[80] ISSD Bilisim Elektronik. *VISSIM: Traffic Simulation*, 2017. [online]. 2017. [Accessed 2018-01-22]. Available from: `http://www.issd.com.tr/en/22992/PTV-VISSIM-Traffic-Simulation-Software`

[81] Jain, S. et al. *Exploiting Mobility for Energy Efficient Data Collection in Wireless Sensor Networks*. Mobile Networks and Applications (MONET), 11(3):327–339, 2006.

[82] Jain, Sushant, Fall, Kevin, and Patra, Rabin. *Routing in a Delay Tolerant Network*. SIGCOMM Comput. Commun. Rev., 34(4):145–158, 2004. ISSN 0146-4833. doi:10.1145/1030194.1015484.

[83] Janeček, J. and Bílý, M. Lokální sítě. Vydavatelství ČVUT Praha, 2004. ISBN 80-01-02900-X.

[84] Jiang, Shan, Ferreira, Joseph, and Gonzalez, Marta C. *Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore*. IEEE Transactions on Big Data, 3(2):208–219, 2017.

[85] Johnson, D.B. et al. *Dynamic Source Routing in Ad Hoc Networks*. Mobile Computing, T. Imielinski and H. Korth, Eds., Kulwer, pp. 152–181, 1996.

[86] Joumaa, C., Caminada, A., and Lamrous, S. *Mask Based Mobility Model A new mobility model with smooth trajectories*. Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks and Workshops, WiOpt 2007, 5th International Symposium, pp. 1–6, 2007.

[87] Kashtan, N, Itzkovitz, S, Milo, R, and Alon, U. *Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs*. Bioinformatics (Oxford, England), 20(11):1746—1758, 2004. ISSN 1367-4803. doi: 10.1093/bioinformatics/bth163.

[88] Kashyap, R.L. and Chellappa, R. *Estimation and Choice of Neighborhoods in Spatial Interaction Models of Images*. Multidimensional Signal Processing, IEEE Press Books, 1986.

[89] Kashyap, R.L. and Khotanzad, A. *A Model Based Method for Rotation Invariant Texture Classification*. IEEE Transaction on Pattern Analysis and Machine Intelligence, 8(4):472–481, 1986.

[90] Kenyon, H.S. *Connectionless Networks Enhance Sensor Efficiency*, 2006. [online]. 2006. [Accessed 2018-01-22]. Available from: `https://www.afcea.org/content/?q=node/1151`

[91] Keranen, A., Ott, J., and Kerkkainen, T. *The ONE Simulator for DTN Protocol Evaluation*. IEEE Communications Magazine, 53(3):38–46, 2015.

[92] Keranen, A. at al. *The ONE - The Opportunistic Network Environment simulator*. Downloaded from https://akeranen.github.io/the-one/, 2015.

[93] Keränen, Ari, Ott, Jörg, and Kärkkäinen, Teemu. *The ONE Simulator for DTN Protocol Evaluation*. In Proceedings of the 2Nd International Conference on Simulation Tools and Techniques, Simutools '09, pp. 55:1–55:10. ICST, Brussels, Belgium, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2009. ISBN 978-963-9799-45-5. doi:10.4108/ICST.SIMUTOOLS2009.5674. Available from: https://doi.org/10.4108/ICST.SIMUTOOLS2009.5674

[94] Keranen, A. et al. *The ONE: The Opportunistic Network Environment simulator*, 2017. [online]. 2017. [Accessed 2018-01-22]. Available from: https://akeranen.github.io/the-one/

[95] Kim, J., Sridhara, V., and Bohacek, S. *Realistic Mobility Simulation of Urban Mesh Networks*, 2007. [online]. 2007. [Accessed 2018-01-22]. Available from: https://www.eecis.udel.edu/~bohacek/Papers/UDelMobility.pdf

[96] Kim, M. et al. *Extracting a Mobility Model from Real User Traces*. Proc. 26th Annual IEEE Conference on Computer Communications (INFOCOM'06), 2006.

[97] Kisilevich, Slava, Mansmann, Florian, Nanni, Mirco, and Rinzivillo, Salvatore. *Spatio-temporal clustering*. In Data mining and knowledge discovery handbook, pp. 855–874. Springer, 2009.

[98] Kleinberg, J. *The Small Worlds Phenomenon: An Algorithmic Perspective*. Proc. of 32nd ACM Symposium on Theory of Computing, 2000.

[99] Le Boudec, J. *Understanding The Simulation of Mobility Models with Palm Calculus.* Technical Report EPFL/IC/2005/033, EPFL, 2005.

[100] Lee, K. et al. *Modeling Steady-state and Transient Behaviors of User Mobility: Formulation, Analysis, and Application.* Proc. 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MOBIHOC'06), 2006.

[101] Legendre, F., Barrel, V., de Amorim, M.D., and Fdida, S. *Reconsidering microscopic mobility modeling for self-organizing networks.* IEEE Network, 20(6):4–12, 2006.

[102] Leguay, J., Friedman, T., and Conan, V. *Evaluating Mobility Pattern Space Routing for DTNs.* Proc. IEEE Infocom 2006, 2006.

[103] Leguay, Jérémie, Friedman, Timur, and Conan, Vania. *DTN routing in a mobility pattern space.* In Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, pp. 276–283. ACM, 2005.

[104] Lember, J. and Koloydenko, A. *The adjusted Viterbi training for hidden Markov models, Bernoulli.* Bernoulli, 14(1):180–206, 2008.

[105] Leong, B., Liskov, B., and Morris, R. *Greedy Virtual Coordinates for Geographic Routing.* In 2007 IEEE International Conference on Network Protocols, pp. 71–80. 2007. ISSN 1092-1648. doi:10.1109/ICNP.2007.4375838.

[106] LEONG, Ben. *Geographic Routing without Planarization.* Proceedings of the 3rd Symposium on Network Systems Design and Implementation (NSDI 2006). San Jose, CA, May, 2006.

[107] Leong, Ben, Liskov, Barbara, and Morris, Robert. *Greedy virtual coordinates for geographic routing.* In 2007 IEEE International Conference on Network Protocols, pp. 71–80. IEEE, 2007.

[108] Li, Feng and Wu, Jie. *LocalCom: a community-based epidemic forwarding scheme in disruption-tolerant networks.* In 2009 6th annual IEEE communications society conference on sensor, mesh and ad hoc communications and networks, pp. 1–9. IEEE, 2009.

[109] Li, Jinyang, Jannotti, John, De Couto, Douglas S. J., Karger, David R., and Morris, Robert. *A Scalable Location Service for Geographic Ad Hoc Routing.* In Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, MobiCom '00, pp. 120–130. New York, NY, USA: ACM, 2000. ISBN 1-58113-197-6. doi:10.1145/345910.345931. Available from: http://doi.acm.org/10.1145/345910.345931

[110] Li, Wenye and Schuurmans, Dale. *Modular community detection in networks.* In Twenty-Second International Joint Conference on Artificial Intelligence. 2011.

[111] Li, Yong, Su, Guolong, Wu, Dapeng Oliver, Jin, Depeng, Su, Li, and Zeng, Lieguang. *The impact of node selfishness on multicasting in delay tolerant networks.* IEEE Transactions on Vehicular Technology, 60(5):2224–2238, 2011.

[112] Lilien, L., Gupta, A., and Yang, Z. *Opportunistic Networks for Emergency Applications and Their Standard Implementation Framework.* International Performance Computing and Communications Conference, IPCCC 2007, pp. 588–593, 2007.

[113] Lin, Y., Li, B., and Liang, B. *Stochastic analysis of network coding in epidemic routing.* Selected Areas in Communications, IEEE Journal, 26(5):794–808, 2008.

[114] Lindgren, Anders, Doria, Avri, and Schelén, Olov. *Probabilistic routing in intermittently connected networks.* In ACM International Symposium on Mobilde Ad Hoc Networking and Computing, MobiHoc 2003: 01/06/2003-03/06/2003. 2003.

[115] LIU, Cong and WU, Jie. *Routing in a cyclic mobispace.* In Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing, pp. 351–360. ACM, 2008.

[116] LIU, Xintao and BAN, Yifang. *Uncovering spatio-temporal cluster patterns using massive floating car data.* ISPRS International Journal of Geo-Information, 2(2):371–384, 2013.

[117] MACKAY, D.J.C. *Information theory, inference, and learning algorithms, Chapter 20: An Example Inference Task: Clustering.* Cambridge University Press, pp. 284–292, 2003.

[118] MADAN, A., CHRONIS, I., and PENTLAND, A. *Mit Social Evolution Dataset.* Downloaded from `http://realitycommons.media.mit.edu/socialevolution.html`, 2008.

[119] MADAN, Anmol, CEBRIAN, Manuel, MOTURU, Sai, FARRAHI, Katayoun, ET AL. *Sensing the" health state" of a community.* IEEE Pervasive Computing, 11(4):36–45, 2012.

[120] MARCHIORI, Massimo and LATORA, Vito. *Harmony in the small-world.* Physica A: Statistical Mechanics and its Applications, 285(3):539 – 546, 2000. ISSN 0378-4371. doi:https://doi.org/10.1016/S0378-4371(00)00311-3.

[121] MASACIOGLU, M. *Control Based Mobile Ad Hoc Networking for survivable, dynamic, mobile Special Operation Force communications.* Calhoun: The NPS Institutional Archive, 2009. [online]. 2009. [Accessed 2018-01-22]. Available from: `http://hdl.handle.net/10945/4592`

[122] MASOUDI-NEJAD, A., SCHREIBER, F., and KASHANI, Z.R.M. *Building blocks of biological networks: a review on major network motif discovery algorithms.* IET Systems Biology, 6:164–174(10), 2012. ISSN 1751-8849.

[123] MERONI, Paolo, GAITO, Sabrina, PAGANI, Elena, and ROSSI, Gian Paolo. *CRAWDAD data set unimi/pmtr (v. 2008-12-01).* Downloaded December, 2008.

[124] MERRIAM-WEBSTER.COM. *Merriam Webster Dictionary*, 2018. [online]. 2018. [Accessed 2018-01-22]. Available from: `http://www.merriam-webster.com/`

[125] MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D., and ALON, U. *Network Motifs: Simple Building Blocks of Complex Networks.* Science, 298(5594):824–827, 2002. ISSN 0036-8075. doi:10.1126/science.298.5594.824.

[126] MILO, R., SHEN-ORR, S., ITZKOVITZ, S., KASHTAN, N., CHKLOVSKII, D., and ALON, U. *Network Motifs: Simple Building Blocks of Complex Networks.* Science, 298(5594):824–827, 2002. ISSN 0036-8075. doi:10.1126/science.298.5594.824.

[127] Milo, Ron, Itzkovitz, Shalev, Kashtan, Nadav, Levitt, Reuven, Shen-Orr, Shai, Ayzenshtat, Inbal, Sheffer, Michal, and Alon, Uri. *Superfamilies of Evolved and Designed Networks*. Science, 303(5663):1538–1542, 2004. ISSN 0036-8075. doi:10.1126/science.1089167.

[128] MIT Human Dynamics Lab. *Reality Commons: Reality Mining Dataset*, 2004. [online]. 2004. [Accessed 2018-01-22]. Available from: `http://realitycommons.media.mit.edu/realitymining.html`

[129] Moreira, Waldir and Mendes, Paulo. *Social-aware opportunistic routing: The new trend*. In Routing in Opportunistic Networks, pp. 27–68. Springer, 2013.

[130] Moreira, Waldir, Mendes, Paulo, and Sargento, Susana. *Assessment model for opportunistic routing*. In 2011 IEEE Third Latin-American Conference on Communications, pp. 1–6. IEEE, 2011.

[131] Moreira, Waldir, Mendes, Paulo, and Sargento, Susana. *Opportunistic routing based on daily routines*. In 2012 IEEE international symposium on a world of wireless, mobile and multimedia networks (WoWMoM), pp. 1–6. IEEE, 2012.

[132] Mousavi, S. M., Rabiee, H. R., Moshref, M., and Dabirmoghaddam, A. *MobiSim: A Framework for Simulation of Mobility Models in Mobile Ad-Hoc Networks*. Wireless and Mobile Computing, Networking and Communications, 2007, WiMOB 2007, 2007.

[133] Musolesi, M., Hailes, S., and Mascolo, C. *Adaptive Routing for Intermittently Connected Mobile Ad Hoc Networks*. Proc. 6th IEEE Int'l. Symp. World of Wireless, Mobile and Multimedia Networks (WoWMoM 2005), 2005.

[134] Musolesi, M. and Mascolo, C. *A Community Based Mobility Model for Ad Hoc Network Research*. In Proc. 2nd ACM/SIGMOBILE International Workshop on Multi-hop Ad Hoc Networks: from theory to reality (REALMAN'06), 2006.

[135] Musolesi, M. and Mascolo, C. *CAR: Context-Aware Adaptive Routing for Delay-Tolerant Mobile Networks*. Mobile Computing, IEEE Transactions, 8(2):246, 2009.

[136] Nain, P. et al. *Properties of Random Direction Models*. INRIA technical report RR-5284, 2004.

[137] National Aeronautics and Space Administration (NASA). *Disruption Tolerant Networking*, 2017. [online]. 2017. [Accessed 2018-01-22]. Available from: `https://www.nasa.gov/directorates/heo/scan/engineering/technology/txt_dtn.html`

[138] National Aeronautics and Space Administration (NASA). *The Transportation Analysis and Simulation System (TRANSIMS)*, 2018. [online]. 2018. [Accessed 2018-01-22]. Available from: `https://code.google.com/archive/p/transims/`

[139] Naumov, V. *Realistic Vehicular Traces*, 2006. [online]. 2006. [Accessed 2018-01-22]. Available from: `http://www.lst.inf.ethz.ch/research/ad-hoc/realistic-vehicular-traces.html`

[140] Nazir, Fawad, Ma, Jianhua, and Seneviratne, Aruna. *Time critical content delivery using predictable patterns in mobile social networks.* In 2009 International Conference on Computational Science and Engineering, vol. 4, pp. 1066–1073. IEEE, 2009.

[141] Neely, M.J. *Capacity and delay tradeoffs for ad-hoc mobile networks.* IEEE Trans. Inf. Theory, 51(6):1917–1937, 2005.

[142] Newman, M. E. J. *Modularity and community structure in networks.* Proceedings of the National Academy of Sciences, 103(23):8577–8582, 2006. ISSN 0027-8424. doi:10.1073/pnas.0601602103.

[143] Newman, M. E. J. and Girvan, M. *Finding and evaluating community structure in networks.* Phys. Rev. E, 69:026113, 2004. doi:10.1103/PhysRevE.69.026113.

[144] Nguyen, H. *A.; Giordano, S.; Spatiotemporal routing algorithm in opportunistic networks.* World of Wireless, Mobile and Multimedia Networks - WoWMoM 2008 International Symposium, pp. 1–6, 2008.

[145] Niu, J., Zhou, X., Wang, K., and Ma, J. *A Data Transmission Scheme for Community-Based Opportunistic Networks.* Wireless Communications, Networking and Mobile Computing, WiCom '09. 5th International Conference, pp. 1–5, 2009.

[146] Palla, Gergely, Derényi, Imre, Farkas, Illés, and Vicsek, Tamás. *Uncovering the overlapping community structure of complex networks in nature and society.* Nature, 435(7043):814 – 818, 2005. ISSN 00280836.

[147] Paramics. *Traffic microsimulation*, 2018. [online]. 2018. [Accessed 2018-01-22]. Available from: `http://www.paramics-online.com/what-is-microsimulation.php`

[148] Park, V. et al. *A Highly Adaptive Distributed Routing Algorithm for Mobile Wireless Networks.* Proceedings of INFOCOM'97, 1997. [online]. 1997. [Accessed 2018-01-22]. Available from: `https://www.comp.nus.edu.sg/~bleong/geographic/related/park97tora.pdf`

[149] Pazand, B. and McDonald, C. *A Critique of Mobility Models for Wireless Network Simulation.* Computer and Information Science, ICIS 2007. 6th IEEE/ACIS International Conference, pp. 141–146, 2007.

[150] Pedace, Jessica, Senatore, Alfonso, and Mendicino, Giuseppe. *Assessing the Reliability of a Fully Coupled Eco-Hydrological Model Based on the Macroscopic Cellular Automata Approach.* Numerical Computations: Theory and Algorithms, p. 109, 2013.

[151] Pelusi, L., Passarella, A., and Conti, M. *Beyond MANETs: Dissertation on Opportunistic Networking.* IITCNR Tech. Rep., 2006.

[152] Pelusi, L., Passarella, A., and Conti, M. *Opportunistic networking: data forwarding in disconnected mobile ad hoc networks.* IEEE: Communications Magazine, Vol. 44(No. 11):134–141, 2006.

[153] Perkins, C. *Ad-Hoc On Demand Distance Vector (AODV) Routing*, 1997. [online]. 1997. [Accessed 2018-01-22]. Available from: `http://www.ics.uci.edu/~atm/adhoc/paper-collection/perkins-draft-ietf-manet-aodv-00.txt`

[154] Perkins, C. et al. *Highly dynamic destination sequenced distance-vector routing (DSDV) for mobile computers.* SIGCOMM 94, pp. 234–244, 1994.

[155] Picard, F., Daudin, J.-J., Koskas, M., Schbath, S., and Robin, S. *Assessing the Exceptionality of Network Motifs.* Journal of Computational Biology, 15(1):1–20, 2008. doi:10.1089/cmb.2007.0137. PMID: 18257674.

[156] Pietilainen, Anna-Kaisa and Diot, Christophe. *CRAWDAD dataset thlab/sigcomm2009 (v. 2012-07-15).* Downloaded from `https://crawdad.org/thlab/sigcomm2009/20120715`, 2012. doi:10.15783/C70P42.

[157] Pietiläinen, Anna-Kaisa, Oliver, Earl, LeBrun, Jason, Varghese, George, and Diot, Christophe. *MobiClique: Middleware for Mobile Social Networking.* In Proceedings of the 2Nd ACM Workshop on Online Social Networks, WOSN '09, pp. 49–54. New York, NY, USA: ACM, 2009. ISBN 978-1-60558-445-4. doi:10.1145/1592665.1592678. Available from: `http://doi.acm.org/10.1145/1592665.1592678`

[158] Pietiläinen, Anna-Kaisa, Oliver, Earl, LeBrun, Jason, Varghese, George, and Diot, Christophe. *MobiClique: Middleware for Mobile Social Networking.* In Proceedings of the 2Nd ACM Workshop on Online Social Networks, WOSN '09, pp. 49–54. New York, NY, USA: ACM, 2009. ISBN 978-1-60558-445-4. doi:10.1145/1592665.1592678. Available from: `http://doi.acm.org/10.1145/1592665.1592678`

[159] Pinjari, Abdul and Bhat, Chandra. *Activity-based travel demand analysis.* A Handbook of Transport Economics, pp. 213–248, 2011.

[160] Poocharoen, P., Magana, M. E., and Alban, E. X. *Partial network coding with cooperation: A cross-layer design for multi-hop wireless networks.* Ultra Modern Telecommunications & Workshops, ICUMT '09, International Conference, pp. 1–8, 2009.

[161] Qiang, Z., Jing, Y., and Minghui, W. *Formal taxonomy research on opportunistic networks.* In Broadband Network & Multimedia Technology, 2009. IC-BNMT'09. 2nd IEEE International Conference on, pp. 854–857. IEEE, 2009.

[162] Rabiner, L.R. *A tutorial on hidden Markov models and selected applications in speech recognition.* Proceedings of the IEEE, 77(2):257–286, 1989.

[163] Raghavan, Usha Nandini, Albert, Réka, and Kumara, Soundar. *Near linear time algorithm to detect community structures in large-scale networks.* Phys. Rev. E, 76:036106, 2007. doi:10.1103/PhysRevE.76.036106.

[164] Rajan, M. A. et al. *Concepts of Graph Theory Relevant to Ad-hoc Networks.* Int. J. of Computers, Communications & Control, Vol. III, 2008. [online]. 2008. [Accessed 2018-01-22]. Available from: `http://www.britannica.com/bps/additionalcontent/18/33373769/Concepts-of-Graph-Theory-Relevant-to-Adhoc-Networks`

[165] Ramasamy, U. *Classification of Self-Organizing Hierarchical Mobile Adhoc Network Routing Protocols - A Summary.* International Journal of Advanced Computer Science and Applications – IJACSA, 1(4), 2010.

[166] Ravikiran, G. and Singh, S. *Influence of mobility models on the performance of routing protocols in ad-hoc wireless networks.* IEEE Vehicular Technology Conference, 4:2185–2189, 2004.

[167] Razzaq, A. and Mehaoua, A. *A robust network coding scheme for SVC- based streaming over wireless mesh network.* Global Information Infrastructure Symposium, GIIS '09, pp. 1–5, 2009.

[168] Reades, J., Calabrese, F., Sevtsuk, A., and Ratti, C. *Cellular Census: Explorations in Urban Data Collection.* IEEE Pervasive Computing, 6(3):30–38, 2007. ISSN 1536-1268. doi:10.1109/MPRV.2007.53.

[169] Rochat, Yannick. *Closeness Centrality Extended to Unconnected Graphs: the Harmonic Centrality Index.* ASNA, Zürich, 2009.

[170] Rojas, A. et al. *Experimental validation of the random waypoint mobility model through a real world mobility trace for large geographical areas.* Proc. 8th ACM international symposium on Modeling, analysis and simulation of wireless and mobile systems, 2005.

[171] Royer, E. M., Melliar-Smith, P. M., and Moser, L. E. *An Analysis of the Optimum Node Density for Ad hoc Mobile Networks.* Proc. of the IEEE International Conference on Communications (ICC), 2001.

[172] Rue, H. and Held, L. Gaussian Markov Random Fields: Theory and Applications. CRC Press, 2005. ISBN 0203492021.

[173] Sarafijanovic-Djukic, Natasa, Pidrkowski, Michal, and Grossglauser, Matthias. *Island hopping: Efficient mobility-assisted forwarding in partitioned networks.* In 2006 3rd Annual IEEE Communications Society on Sensor and Ad Hoc Communications and Networks, vol. 1, pp. 226–235. IEEE, 2006.

[174] Schneider, Christian M., Belik, Vitaly, Couronné, Thomas, Smoreda, Zbigniew, and Gonzáles, Marta C. *Unravelling daily human mobility motifs*. Journal of the Royal Society Interface, 10(84):20130246, 2013. doi:10.1098/rsif.2013.0246.

[175] Schneider, Christian M, Belik, Vitaly, Couronné, Thomas, Smoreda, Zbigniew, and González, Marta C. *Unravelling daily human mobility motifs*. Journal of The Royal Society Interface, 10(84):20130246, 2013.

[176] Schreiber, Falk and Schwöbbermeyer, Henning. *Frequency Concepts and Pattern Detection for the Analysis of Motifs in Networks*. In Corrado Priami, Emanuela Merelli, Pablo Gonzalez, and Andrea Omicini, editors, Transactions on Computational Systems Biology III, pp. 89–104. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005. ISBN 978-3-540-31446-2.

[177] Schreiber, Falk and Schwöbbermeyer, Henning. *MAVisto: a tool for the exploration of network motifs*. Bioinformatics, 21(17):3572–3574, 2005. ISSN 1367-4803. doi:10.1093/bioinformatics/bti556.

[178] Schwöbbermeyer, H. *Network Motifs*. In Bjrn H. Junker and Falk Schreiber, editors, Analysis of Biological Networks (Wiley Series in Bioinformatics), pp. 85–108. New York, NY, USA: Wiley-Interscience, 2008. ISBN 0470041447, 9780470041444.

[179] Scott, James, Gass, Richard, Crowcroft, Jon, Hui, Pan, Diot, Christophe, and Chaintreau, Augustin. *CRAWDAD dataset cambridge/haggle (v. 2006-09-15)*. Downloaded from `https://crawdad.org/cambridge/haggle/20060915/imote`, 2006. doi:10.15783/C73S3N. Traceset: imote.

[180] Seferoglu, H. and Markopoulou, A. *Opportunistic network coding for video streaming over wireless network*. Packet Video 2007, pp. 191–200, 2007.

[181] Seferoglu, H. and Markopoulou, A. *Video-aware opportunistic network coding over wireless networks*. Selected Areas in Communications, IEEE Journal, 27(5):713–728, 2009.

[182] Sharma, G., Mazumdar, R., and Shroff, B. *Delay and Capacity Trade-Offs in Mobile Ad Hoc Networks: A Global Perspective Networking*. IEEE/ACM Transactions, 15(5):981–992, 2007.

[183] Sharma, G. and R. *Mazumdar. Scaling laws for capacity and delay in wireless ad hoc networks with random mobility*. Proc. IEEE ICC, pp. 3869–3873, 2004.

[184] Socievole, Annalisa, De Rango, Floriano, and Caputo, A. *Opportunistic mobile social networks: From mobility and Facebook friendships to structural analysis of user social behavior*. Computer Communications, 87:1–18, 2016.

[185] SOCIEVOLE, Annalisa, DE RANGO, Floriano, and COSCARELLA, Carmine. *Routing approaches and performance evaluation in delay tolerant networks.* In 2011 Wireless Telecommunications Symposium (WTS), pp. 1–6. IEEE, 2011.

[186] SOCIEVOLE, Annalisa, YONEKI, Eiko, DE RANGO, Floriano, and CROWCROFT, Jon. *ML-SOR: Message routing using multi-layer social networks in opportunistic communications.* Computer Networks, 81:201–219, 2015.

[187] SOCIOPATTERNS.ORG. *SocioPatterns*, 2019. [online]. 2019. [Accessed 2019-03-29]. Available from: `http://www.sociopatterns.org/gallery/`

[188] SPACCAPIETRA, Stefano, PARENT, Christine, DAMIANI, Maria Luisa, DE MACEDO, Jose Antonio, PORTO, Fabio, and VANGENOT, Christelle. *A conceptual view on trajectories.* Data & knowledge engineering, 65(1):126–146, 2008.

[189] SPAHO, Evjola, BYLYKBASHI, Kevin, BAROLLI, Leonard, KOLICI, Vladi, and LALA, Algenti. *Evaluation of different DTN routing protocols in an opportunistic network considering many-to-one communication scenario.* In 2016 19th International Conference on Network-Based Information Systems (NBiS), pp. 64–69. IEEE, 2016.

[190] SPYROPOULOS, T., PSOUNIS, K., and RAGHAVENDRA, C. S. *Efficient Routing in Intermittently Connected Mobile Networks: The Single-Copy Case.* IEEE/ACM Transactions on Networking, 16(1):63–76, 2008. ISSN 1063-6692. doi:10.1109/TNET.2007.897962.

[191] SPYROPOULOS, Thrasyvoulos, PSOUNIS, Konstantinos, and RAGHAVENDRA, Cauligi S. *Spray and wait: an efficient routing scheme for intermittently connected mobile networks.* In Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking, pp. 252–259. ACM, 2005.

[192] SPYROPOULOS, Thrasyvoulos, PSOUNIS, Konstantinos, and RAGHAVENDRA, Cauligi S. *Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility.* In Fifth Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerComW'07), pp. 79–85. IEEE, 2007.

[193] SPYROPOULOS, Thrasyvoulos, PSOUNIS, Konstantinos, and RAGHAVENDRA, Cauligi S. *Efficient routing in intermittently connected mobile networks: The multiple-copy case.* IEEE/ACM Transactions on Networking (ToN), 16(1):77–90, 2008.

[194] SPYROPOULOS, Thrasyvoulos, RAIS, Rao Naveed, TURLETTI, Thierry, OBRACZKA, Katia, and VASILAKOS, Athanasios. *Routing for Disruption Tolerant Networks: Taxonomy and Design.* Wirel. Netw., 16(8):2349–2370, 2010. ISSN 1022-0038. doi:10.1007/s11276-010-0276-9.

[195] Srinivasan, V. et al. *Analysis and Implications of Student Contact Patterns Derived from Campus Schedules.* Proc. The 12th ACM International Conference on Mobile Computing and Networking (MOBICOM'06), 2006.

[196] Stepanov, I. *Mobility Modeling of Outdoor Scenarios for MANETs.* Proceedings of the 38th Annual Simulation Symposium (ANSS'05), pp. 312–322, 2005.

[197] Tang, John Kit, LEONTIADIS, L, SCELLATO, S, et al. Temporal network metrics and their application to real world networks. Ph.D. thesis, University of Cambridge, 2012.

[198] Theodoridis, S. and Koutroumbas, K. Pattern Recognition. Academic Press, 2009. ISBN 978-1-59749-272-0.

[199] Timo, R., Blackmore, K., and Hanlen, L. *Strong Stochastic Stability for MANET Mobility Models.* Networks, 2007. ICON 2007. 15th IEEE International Conference, pp. 13–18, 2007.

[200] Toch, Eran, Lerner, Boaz, Ben-Zion, Eyal, and Ben-Gal, Irad. *Analyzing large-scale human mobility data: a survey of machine learning methods and applications.* Knowledge and Information Systems, 58:501–523, 2018.

[201] Toumpis, S. *Large wireless networks under fading, mobility, and delay constraints.* Proc. IEEE INFOCOM, pp. 609–619, 2004.

[202] Tower, J.P. and Little, T.D.C. *A Proposed Scheme for Epidemic Routing with Active Curing for Opportunistic Networks.* Advanced Information Networking and Applications - Workshops, AINAW 2008. 22nd International Conference, pp. 1696–1701, 2008.

[203] Tuceryan, M. and K., Jain A. *Texture Analysis, The Handbook of Pattern Recognition and Computer Vision (2nd Edition).* World Scientific Publishing Co., pp. 207–248, 1998.

[204] Tuduce, C. et al. *A Mobility Model Based on WLAN Traces and its Validation.* Proc. 25th Annual IEEE Conference on Computer Communications (INFOCOM'05), 2005.

[205] University of Florida. *CORSIM: Microscopic Traffic Simulation Model,* 2006. [online]. 2006. [Accessed 2018-01-22]. Available from: `http://www-mctrans.ce.ufl.edu/featured/TSIS/Version5/corsim.htm`

[206] Vahdat, Amin, Becker, David, et al. *Epidemic routing for partially connected ad hoc networks.* Technical Report CS-200006, 2000.

[207] Vapnik, V.N. The Nature of Statistical Learning Theory. Springer-Verlag, 1995. ISBN 0-387-98780-0.

[208] Vapnik, V.N. and Kotz, S. Estimation of Dependences Based on Empirical Data. Springer, 2006. ISBN 0-387-30865-2.

[209] Verma, A. and Srivastava, A. *Integrated Routing protocols for Opportunistic Networks*. (IJACSA) International Journal of Advanced Computer Science and Applications, 2(3), 2011.

[210] Wang, H., Chang, R.Y., and Kuo, C.-C.J. *Wireless Multi-party video conferencing with network coding, Multimedia and Expo.* ICME 2009 - IEEE International Conference, pp. 1492–1495, 2009.

[211] Wang, Y. et al. *Protocol Design and Optimization for Delay/Fault-Tolerant Mobile Sensor Networks*. Distributed Computing Systems, ICDCS '07. 27th International Conference, p. 7, 2007.

[212] Watts, D. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, 1999.

[213] Wernicke, Sebastian. *Efficient detection of network motifs*. IEEE/ACM transactions on computational biology and bioinformatics, 3(4):347—359, 2006. ISSN 1545-5963. doi:10.1109/tcbb.2006.51.

[214] Widhalm, Peter, Yang, Yingxiang, Ulm, Michael, Athavale, Shounak, and González, Marta C. *Discovering urban activity patterns in cell phone data*. Transportation, 42(4):597–623, 2015. ISSN 1572-9435. doi:10.1007/s11116-015-9598-x.

[215] Widmer, J. and Le Boudec, Y. *Network Coding for Efficient Communication in Extreme Networks*. Proc. ACM SIGCOMM 2005 Wksp. Delay Tolerant Networks, Philadelphia, PA, 2005.

[216] Wikipedia. *Ad (latina)*, 2017. [online]. 2017. [Accessed 2018-01-22]. Available from: `http://cs.wikipedia.org/wiki/Ad_%28latina%29`

[217] Wikipedia. *Support vector machine*, 2017. [online]. 2017. [Accessed 2018-01-22]. Available from: `https://en.wikipedia.org/wiki/Support_vector_machine`

[218] Wikipedia. *Cluster analysis*, 2018. [online]. 2018. [Accessed 2018-01-22]. Available from: `https://en.wikipedia.org/wiki/Cluster_analysis`

[219] Woungang, I., Dhurandher, S.K., Anpalagan, A., and Vasilakos, A.V. Routing in Opportunistic Networks. SpringerLink : Bücher. Springer New York, 2013. ISBN 9781461435143. Available from: `https://books.google.cz/books?id=mMYBTCBDtAsC`

[220] Woungang, Isaac, Dhurandher, Sanjay Kumar, Anpalagan, Alagan, and Vasilakos, Athanasios V. Routing in opportunistic networks. Springer, 2013.

[221] Xia, Feng, Liu, Li, Li, Jie, Ma, Jianhua, and Vasilakos, Athanasios V. *Socially aware networking: A survey.* IEEE Systems Journal, 9(3):904–921, 2015.

[222] Yan, Zhixian, Parent, Christine, Spaccapietra, Stefano, and Chakraborty, Dipanjan. *A hybrid model and computing platform for spatio-semantic trajectories.* In Extended Semantic Web Conference, pp. 60–75. Springer, 2010.

[223] Yang Hao, Jiang, Hatzopoulou, Marianne, and J. Miller, Eric. *Integrating an Activity-Based Travel Demand Model with Dynamic Traffic Assignment and Emission Models.* Transportation Research Record: Journal of the Transportation Research Board, 2176:1–13, 2010. doi:10.3141/2176-01.

[224] Yoon, J., Liu, M., and Noble, B. *Random Waypoint Considered Harmful.* Proc of Infocom, 2003.

[225] Yoon, J. et al. *Building realistic mobility models from coarse-grained traces.* Proc. ACM International Conference On Mobile Systems, Applications And Services (MobiSys'06), pp. 177–190, 2006.

[226] Yu, C.-M., Lu, C.-S., and Kuo, S.-Y. *Habitual Behavior-Based Opportunistic Data Forwarding in Wildlife Tracking Wireless Communication Systems.* ISWCS 2007. 4th International Symposium, pp. 807–808, 2007.

[227] Zatko, P. *Military Networking Protocol (MNP)*, 2012. [online]. 2012. [Accessed 2018-01-22]. Available from: `http://archive.is/d9Tz#selection-446.1-473.34`

[228] Zhang, Z. *Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges.* IEEE Communications Surveys Tutorials, 8(1):24–37, 2006. ISSN 1553-877X. doi:10.1109/COMST.2006.323440.

[229] Zhao, W., Ammar, M., and Zegura, E. *A Message Ferrying Approach for Data Delivery in Sparse Mobile Ad Hoc Networks.* Proc. 5th ACM Int'l. Symp. Mobile Ad Hoc Networking and Computing (Mobihoc), ACM Press, pp. 187–198, 2004.

[230] Zheng, Q. et al. *An Agenda-based Mobility Model.* Proc. 39th IEEE Annual Simulation Symposium (ANSS-39-2006), Huntsville AL, 2006.

[231] Zhu, H. and Lu, K. *Resilient Opportunistic Forwarding: Issues and Challenges.* Military Communications Conference, MILCOM 2007, IEEE, pp. 1–7, 2007.

Reviewed Publications of the Author Relevant to the Thesis

[A.1] Smítková Janků, L. and Hyniová, K. Improvement of Routing in Opportunistic Communication Networks of Vehicles by Unsupervised Machine Learning EANN 2019, 20th International Conference on Engineering Application of Neural Networks (www.eann2019.eu) Crete, Greece. Proceedings of EANN 2019, Publisher: Springer, Communications in Computer and Information Science Series 2019 (accepted paper)

[A.2] Smítková Janků, L. and Hyniová, K. Enhanced Routing Algorithm for Opportunistic Networking: On the Improvement of the Basic Opportunistic Networking Routing Algorithm by the Application of Machine Learning. In: Recent Researches in Information Science & Applications: ICPRAM 2014, International Conference on Pattern Recognition Applications and Methods (www.icpram.org) Angers, France. Proceedings of ICPRAM2014, Publisher: SCITEPRESS, selected papers appear in Proceedings in LNCS Springer-Verlag. 2014 ISBN ISBN 978-1-61804-150-0, ISSN 1790-5109

[A.3] Smítková Janků, L. and Hyniová, K. Communication in Opportunistic Networks: A Proposal of a Mobility Model Based on Texture Analysis using Gaussian Markov Random FieldsNode Mobility Model for Opportunistic Networks based on Gaussian Markov Random Fields and Underlying Hidden Markov Chain. In: Recent Researches in Information Science & Applications: Proceedings of the 7th WSEAS International Conference on COMPUTER ENGINEERING and APPLICATIONS (CEA'13) 2013 ISBN ISBN 978-1-61804-150-0, ISSN 1790-5109

Remaining Publications of the Author Relevant to the Thesis

[A.4] Smítková Janků, L. *IMPROVEMENT OF THE SELECTED ROUTING PROTOCOLS IN OPPORTUNISTIC NETWORKS BY THE APPLICATION OF PATTERN RECOGNITION AND MACHINE LEARNING TECHNIQUES*. Ph.D. Minimum Thesis, Faculty of Information Technology, Prague, Czech Republic, 2014.

[A.5] Smítková Janků, L. Communication in Opportunistic Networks: A Proposal of a Mobility Model Based on Texture Analysis using Gaussian Markov Random Fields. In Proceedings of POSTER 2011: 15th International Student Conference on Electrical Engineering. CTU, Faculty of Electrical Engineering Prague, Czech Republic2011 2011 ISBN 978-80-01-04806-1

[A.6] Janků, L. Computational Auditory Scene Analysis. In: Proceedings of Workshop 2002. Prague, A,B:228-229. 2002

[A.7] Janků, L. and Wagenknecht, M. and Eck, V. and Gocht, U. Technical Systems Safety Analysis and Assessment. In: Proceedings of Workshop 2002. Prague, A,B:228-229. 2002

[A.8] Janků, L. and Eck, V. and Šorf., M. and Macek, J. and Dobiáš, L. and Salač, M. and Fejtová, M. and Choura, M. Towarda the Artificial Life and Brain Sceinces: New Trends in Bionics Edutation. In proceedings of Workshop 2002, I:892-893. 2002

[A.9] Janků, L. An Efficient Machine Listening Approach to Sound Source Separation. In: Proceedings of Workshop 2001. Prague, A:262-263. 2001

[A.10] Janků, L. and Šorf, M. and Eck, V. Preliminary Phase of Co-operation between Human Operator and Intelligent Interface-Computational Personality Type Classification. Praha, CVUT FEL. 2000

Remaining Publications of the Author

[A.11] Hyniová, K. and Smítková Janků, L. Using Fuzzy Logic to Control an Innovative Active Vehicle Suspension System In: Proceedings of the 18 th International Conference on systems. Athens: WSEAS, 2014, pp. 161-167. Recent Advances in Electrical Engineering Series. 2014 ISBN ISBN 978-1-61804-243-9, ISSN 1790-5117

[A.12] Hyniová, K. and Krajl, M. and Smítková Janků, L. Experiments Taken on Energy Management in Active Suspension of Vehicles International Journal of Circuits, Systems and Signal Processing. 2012, 6(1), 196-203. 2012 ISSN 1998-4464

[A.13] Hyniová, K. and Smítková Janků, L. and Honců, J. and Stříbrský, A. On Experiments Taken on The Active Shock Absorber Test Stand In: APPLIED MATHEMATICS in ELECTRICAL and COMPUTER ENGINEERING. New York: WSEAS Press, 2012, pp. 104-109. 2012 ISBN 978-1-61804-064-0

[A.14] Hyniová, K. and Smítková Janků, L. Using Fuzzy Logic to Control an Innovative Active Vehicle Suspension System. In: Proceedings of the 18 th International Conference on systems. 18th international Conference on Systems. Fira, 17.07.2014 - 21.07.2014. Athens: WSEAS. 2014, s. 161-167. Recent Advances in Electrical Engineering Series. ISSN 1790-5117. ISBN 978-1-61804-243-9.

[A.15] Smítková Janků and L. and Hyniová, K. Enhanced routing algorithm for opportunistic networking: On the improvement of the basic opportunistic networking routing algorithm by the application of machine learning. In: ICPRAM 2014 - Proceedings of the 3rd International Conference on Pattern Recognition Applications and Methods. The 3rd International Conference on Pattern Recognition Applications and Methods. Loire Valley; France, 06.03.2014 - 08.03.2014. Paris: Academic Science Research. 2014, s. 771-776. ISBN 978-989-758-018-5. 2014

[A.16] Hyniová, K. et al. On Experiments Taken on The Active Shock Absorber Test Stand. In: APPLIED MATHEMATICS in ELECTRICAL and COMPUTER ENGINEERING. CIRCUITS, SYSTEMS, SIGNAL and TELECOMMUNICATIONS.

Cambridge, Harvard, 25.01.2012 - 27.01.2012. New York: WSEAS Press. 2012, s. 104-109. ISBN 978-1-61804-064-0.

[A.17] Hyniová, K. and Krajl, M. and Smítková Janků, L. Experiments Taken on Energy Management in Active Suspension of Vehicles [online]. International Journal of Circuits, Systems and Signal Processing. 2012, 6(1), s. 196-203. ISSN 1998-4464. Dostupné z: http://www.naun.org/wseas/cms.action?id=3113

[A.18] Smítková Janků, L. and Hyniová, K. Application of Feed-Forward Neural Network and MMI-supervised Vector Quantizer to the Task of Content Based Audio Segmentation by Co-operative Unmanned Flying Robots. In: AL-DABASS, D., PANTELOUS, A., a TAWFIK, H., eds. Proceedings on First International Conference on Intelligent Systems, Modelling and Simulation 2010. First International Conference on Intelligent Systems, Modelling and Simulation. Liverpool, 27.01.2010 - 29.01.2010. Los Alamitos: IEEE Computer Society. 2010, s. 111-115. ISBN 978-1-4244-5984-1.

[A.19] Smítková Janků, L. Auditory Scene Analysis via Application of ICA in a Time-Frequency Domain. In: SOJKA, P., KOPEČEK, I., a PALA, K., eds. Text, Speech and Dialogue, 7th International Conference TSD 2004. 7th International Conference Text, Speech and Dialogue 2004. Brno, 08.09.2004 - 11.09.2004. Berlin: Springer. 2004, s. 347-353. ISSN 0302-9743. ISBN 3-540-23049-1.

[A.20] Smítková Janků, L. and Černocký, J. Content-based General Audio Segmentation. 21.06.2004.

[A.21] Smítková Janků, L. A study of the separability of enviromental sounds in mixtures with respect to their acoustical structures. In: DARWIN, C.J., ed. Auditory Scene Analysis and Speech Perception by Human and Machine 2004. International Conference on Auditory Scene Analysis and Speech Perception by Human and Machine. Delmenhorst, 22.08.2004 - 24.08.2004. Brighton: University of Sussex. 2004, s. 34.

[A.22] Smítková Janků, L. and Smítka, J. Construction of Fully-Connected Hidden Markov Models of Environmental Sounds. In: HAMAM, Y., ed. Proceedings of the 5th EUROSIM Congres Modelling and Simulation. 5th EUROSIM Congress on Modelling and Simulation. Paris, 06.09.2004 - 10.09.2004. Vienna: EUROSIM-FRANCOSIM-ARGESIM. 2004, ISBN 3-901608-28-1.

[A.23] Eck, V. and Křemen, V. and Smítková Janků, L. Komunikace mezi univerzitami a nejen mezi nimi. In: Studienkolleg Sonderheft. Kommunikationsprozesse in Euroregionen. Ústí nad Labem, 24.03.2004. Ústí nad Labem: Univerzita J. E. Purkyně. 2004, s. 17-30.

[A.24] Janků, L. Artificial Perception: Auditory Decomposition of Mixtures of Environmental Sounds - Combining Information Theoretical and Supervised Pattern Recognition Approaches. In Theory and Practice of Computer Science, Lecture Notes in Computer Science., Springer 1:229-240. 2004

[A.25] Janků, L., and Černocký, J. Content-based General Audio Segmentation. In MLMI'04. 2004

[A.26] Janků, L. Modelling Natural Sound Textures. In Proceedings of 38th International Conference MOSIS '04, 237-244. 2004

[A.27] Janků, L. and Lhotská, L. Towards the Applications of Fuzzy Logic to the Sound Recordings Fingerprint Computation and Comparison. In: Recent Advances in Computers, Computing and Communications. New Jersey : WSEAS Press, 271-274. 2002

[A.28] Janků, L. and Křemen, V. and Eck, V. and Lhotská, L. and Dostál, J. Accident Analysis and Safety Assesment in Industrial System Control: Simulation of the Human Supervisory Operator. In: Inteligentní metody pro prumyslovou praxi: Diagnostika, [CD-ROM]. Prague : CTU, Faculty of Electrical Engineering, Department of Cybernetics. 2002

[A.29] Janků, L. and Lhotská, L. and Křemen, V. Additional Method for the Evaluation of the Training Capturing: Intelligent System for the Estimation of the Human Supervisory Operator Stress Level under the Normal Safety Critical Circumstances. In: Intelligent E-Health Applications in Medicine and Other Medical Applications of Adaptive/Hybrid Intelligent Systems. Aegean: University of the Aegean, 1:147-153. 2002

[A.30] Janků, L. Sound Source Separation through Models of Auditory Processes and Fuzzy-Rule System. In: MOSIS 01-Modelling and Simulation of Systems., 1:195-200. 2001

[A.31] Janků, L., and Eck, V. Fuzzy Distributed Reasoning about Multi-Dimensional Medical Data. In: Proceedings: MEDICON 2001. Zagreb, 1:441-444. 2001

[A.32] Janků, L., and Lhotská, L. Aggregation of Fuzzy Preferences as an Efficient Approach to the Computational Personality Type Classification Problem. In: Problems in Applied Mathematics and Computational Intelligence, New Jersey: WSES Press. 2001

[A.33] Janků, L., and Lhotská, L. Musical Instrument Classification through the Model of Auditory Periphery and a Neural Network. In: Advances in Scientific Computing, Computational Intelligence and Applications, New Jersey: WSES Press. 2001

[A.34] Janků, L., and Lhotská, L. Aggregation of Fuzzy Preferences as an Efficient Approach to the Computational Personality Type Classification Problem. In: Proceedings of WSES international conferences NNA 2001, FSFS 2001, EC 2001, Puerto De La Cruz, Canary Islands, Spain. 2001

[A.35] Janků, L., and Lhotská, L. Musical Instrument Classification through the Model of Auditory Periphery and a Neural Network. In: Proceedings of CSCC 2001, MCP 2001, MCME, 2001, IEEE Computer Society Press. 2001

[A.36] Janků, L. and Šorf, M. and Lhotská, L. and Eck, V. Five Approaches to Computational Personality Classification Problem. In: Proceedings of CIMA 2001, ICSC/NAISO conference, UK. 2001

[A.37] Janků, L. and Šorf, M. and Lhotská, L., and Eck, V. Computational Personality Type Classification. In: Proceedings of ICCC 2001. Krakowie : Wydzial Inzynierii Mechanicznej i Robotyki, 1:301-306. 2001

[A.38] Janků L. and Šorf, M. and Eck, V., Lhotská, L., and Fejtová, M. Multiple Criteria Choice Mechanism for Intelligent Man-Machine Interface. In: Proceedings of INES. IEEE Industrial Electronics Society. 2000