Information Dissemination in Vehicular Networks

Tamás Zoltán Veres

Master of Science Thesis

Wireless and Mobile Communications Group (WMC)
Department of Telecommunications
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology

Delft University of Technology

Wireless and Mobile Communications

EQN 2010 03

Information Dissemination in Vehicular Networks

Master of Science Thesis

For the degree of Master of Science in
Wireless and Mobile Communications Group (WMC)
at Department of Telecommunications
at Delft University of Technology

Tamás Zoltán Veres

May 31, 2011

Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Delft, The Netherlands

The work in this thesis was supported by the IDEA League student grant. Their cooperation is hereby gratefully acknowledged.

All rights reserved.
Copyright © Telecommunications Department
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Delft, The Netherlands

DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF
TELECOMMUNICATIONS

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled

INFORMATION DISSEMINATION IN VEHICULAR NETWORKS

by

TAMÁS ZOLTÁN VERES

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE.

Dated: May 31, 2011

Supervisor: ______________________
Dr.Ertan Onur

Readers: ______________________
Prof.dr.ir.Ignas Niemegeers

____
Dr.Christian Doerr

Abstract

A Vehicular Ad-Hoc Network (VANET) is a network where vehicles can disseminate information such as traffic conditions, availability of free parking lots or content. In order to fulfill VANET application requirements (Time To Live (TTL), information distance spread), we need to use the available network resources (e.g., radio capacity) optimally depending on the traffic conditions. While cellular-based approaches (e.g., 3rd Generation mobile telecommunications (3G)) is suited for low traffic conditions, dense traffic (where users might already use 3G) is particularly suited for relying on Vehicle-to-Vehicle (V2V) communication (e.g., ad-hoc WIFI). There is a clear need to adapt the dissemination strategy to the current traffic conditions and application requirements. The optimal solution is based on combining or alternating between cellular-based and opportunistic dissemination strategies.

In this thesis, we aim at modeling epidemic dissemination in vehicular networks for prediction and optimization purposes. Contrarily to state-of-the-art works (e.g., Susceptible, Infectious (SI) models; VANET) we consider the dissemination as a function of time, but also distance. To do so, we first abstract vehicular mobility with a lattice grid and homogeneous transition probabilities from which we derive a dissemination model (with upper and lower bounds). We evaluate our model numerically and find a transition threshold also known as disease outbreak probability. In a next step, we depart from this abstract lattice model to Manhattan mobility and show by simulations that a similar behavior is observed hence validating the analytical model. We further show that our lattice model is also a good approximation of realistic vehicular mobility. Based, on these results we can illustrate how vehicles can adopt the best dissemination strategy (e.g., V2V, 3G or a combination) depending on the traffic conditions (sparse vs. dense) and application requirements (TTL, distance). Our work is a major contribution as it provides a model of spreading which accounts for distance with direct applications to vehicular networks for opti-

mizing dissemination strategies depending on the traffic conditions and application requirements.

Table of Contents

List of Figures

List of Tables

Preface

The idea of doing my thesis on a subject related to vehicles was clear from the beginning of my studies though my deep interest of cars and automotive topics. The lecture taught by my future supervisor at TUDelft made me sure how to combine my interest about vehicles to telecommunications and led me to do special and extensive research related to Vehicular Ad-Hoc Networks.

After finding the way to cooperate with ETH Zürich I could focus on the state-of-art researches and formulate a challenging problem which has not been studied before. Throughout this research work, I was facing several choices of how to get to the next step ahead and I am glad that I always got great help to do so and archive the best I could.

Acknowledgements

It is a pleasure to thank those who made this thesis possible. They have made their support available number of ways and greatly contributed to this work.

I would like to thank Dr.Franck Legendre for his supervision during the writing of this thesis. He has supported me throughout my thesis work with his patience and professional knowledge.

This thesis would not have been possible unless the support of Dr.Ertan Onur who introduced me the topic of this work and encouraged me to work on it abroad in cooperation with ETH Zürich, Switzerland.

I would like to show my gratitude to Prof.Bernhard Plattner to the opportunity to join the research group lead by him.

I am indebted to many of my colleagues and friends to support me throughout my studies at The Netherlands and Switzerland.

I owe my deepest gratitude to my Parents and my Family for their continuous support and love. "Nélkületek nem sikerülhetett volna. Köszönöm!"

Chapter 1

Introduction

Vehicular Ad-Hoc Network (VANET) is going to be briefly introduced in order to position the topic of this work. We show the need for an analytical dissemination model as well as the need to study dissemination strategy's tradeoff for *vehicular networks*.

1-1 About MANET and VANET

VANET comprises vehicles and road side units which are forming a communication network. High level of mobility and a relative short transmission range (100 to 300 meter while for example the length of a road on average or size of a city is higher) implies that not all the nodes are being directly connected. Depending on their density they are either single entities or forming non-connected clusters or all nodes are being connected to each other. The way of communication of nodes is either direct or multi-hop connection. VANET is a special type of Mobile Ad-Hoc Network (MANET). Nodes of such networks are capable of setting up and maintaining a communication network by themselves. They are sharing radio capacity to communicate within their transmission range and relay messages of other communicating parties (multihop communication). Important to mention that in contrast to MANET's vehicles of VANETs are traveling on the existing road network and not moving randomly as in case of MANET's. Vehicles mobility are constrained by the road network/topology, traffic signs and traffic conditions.

VANET utilize various communication technologies in order to meet the VANET application's requirements (e.g., service area, data speed, maximum radio capacity, delay in time). These requirements are not independent from each other and being fundamentally and regulatory limited.

1-2 Motivation

The increasing number of vehicles on road networks are resulting in a higher chance of collision and existence of traffic jams. In order to avoid those cases, VANET-based applications are aiming to reduce the negative effect of the continuous increase of number of vehicles by allowing the vehicles to communicate with each other. The possibility of communication among vehicles and between vehicles and infrastructure creates opportunity for safety and non-safety applications to increase the level of their service by introducing new functions. Disseminated information can range from time-critical accident avoidance information through road condition information to entertainment and advertisement information.

In *vehicular networks* communication can be based on ad-hoc communication technology for vehicles or centralized, infrastructure based (like 3G) technology. *Vehicle hybrid networks* are employing communication technologies of both types. Dissemination strategy of *vehicle hybrid networks* describes which technology and what form need to be employed for optimal information dissemination.

Opportunistic epidemic dissemination is particularly suitable for dense scenarios with high traffic conditions due to the high connectedness of vehicles. However, ad-hoc communication technology-based epidemic dissemination does not provide necessary level of service in case of sparse networks. Therefore, for low traffic conditions infrastructure- and/or Delay Tolerant Network (DTN)-based approaches are necessary (in DTNs the information is usually not time-critical). DTN's carry-and-forward approach is suitable for non-time critical *vehicular applications* while time-critical applications need to employ centralized, infrastructure based (although 3G-based dissemination it is highly resource- and cost-sensitive). Therefore, it is essential to optimize dissemination strategy accounting for *vehicular application's* requirements (e.g., time, distance) and traffic conditions.

Message dissemination in VANETs is being widely studied from the practical point of view which is instead being analyzed as an analytical problem in our work. Similar analytical models from epidemiology, biology and physics accounts for time and rate of infection. In contrast, *vehicular applications* have also requirements on the service area based on how crucial is a message at a certain distance from the source of information. Moreover, vehicular information dissemination does not need to account for birth-death processes, because we assume all vehicles are participating in the information spreading – no birth – and number of vehicles is constant – no birth and death. However, vehicles are traveling as independent entities on the existing road network and following vehicular movement patterns.

1-3 Problem Definition

Vehicles in our study are forming a VANET and they are using single link communication to transfer messages to each other within their communication range. Therefore, as a group of nodes they are capable of disseminating information. Information is going to be transfered to nodes within a certain distance and time limit. Therefore, the dissemination model for information spreading needs to account for time and distance as well.

Considering a single source information of the problem would be to answer the following questions:

- How long does it take to send information to a certain distance?
- How many nodes are going to be aware of the message after certain time?
- How does traffic and the dissemination strategy affects the previously stated questions?

The first question aims to answer, based on the distance, the delay of information spreading while the second question answers the percentage of nodes aware of information for a given time limit. Dissemination in VANET is highly influenced by the current traffic conditions and dissemination strategy employed. Therefore, the third concern is to analyze the tradeoff between *vehicular application's* requirements in connection with the dissemination strategy applied.

Researches in VANET's are mainly based on simulation's results, because real-world testbeds are impracticable due to the large number of vehicles required. Therefore, an analytical model of the problem is necessary to understand information dissemination accounting for distance and time. Such a model would assist as a starting point for creating powerful applications with focusing on the service instead of the network behind their system. Moreover, mimicking realistic movement of vehicles moving on the road network implies that simulations are being highly complex and resource consuming compared to an analytical model. The complexity is more significant if we consider studies for greater distances though the problem of message dissemination scales with the square of distance.

1-4 Objectives

We aim, at finding an analytical model for information spreading among vehicles and gain more insight on resources like time and capacity needed. We validate our analytical model with simulations and study these results by comparing it to different level of realistic simulations performed using a realistic vehicular simulator.

Results of this work are expected to help to understand the tradeoff between requirements of *vehicular applications* (e.g., information awareness rate, time limit, distance) and find the optimum. Moreover, our study is aiming to define a general dissemination strategy employing hybrid technology in order to meet requirements in a resource efficient manner. We believe that our study will reduce the need for early stage real-life experiments by utilizing simulations and analytical studies.

1-5 Milestones of Work and Structure of Thesis

The remainder of this thesis is organized as follows. In Chapter 2, related work will be discussed. The main topic in Chapter 3 will be a simple analytical model which we are going to study in order to find an analytical representation of the problem. We will evaluate our findings with numerical simulations and present limitations of the models. Chapter 4 deals with a vehicular simulator application which functionalities we have extended and implemented in order to support simulations of epidemic type of information dissemination. At first step, we do simulations with the Manhattan Mobility and study the similarities of it to the numerical simulations based on the simple model. Next, we will compare results of the Manhattan Mobility based simulation and two realistic models based on real-world road structure and either employing: (i) interaction-free vehicle movements model, or (ii) realistic vehicular mobility. After the principles and data collection through simulations we shall then go on to study the dissemination tradeoff. We will present requirements of *vehicular applications* through case studies and conclude by proposing a general dissemination strategy for *hybrid vehicular networks*. Eventually, in Chapter 6 we summarize our work and list useful suggestions for further work.

Chapter 2

Literature Review

Researchers working on related topics to vehicular networks form the VANET community. They hold annual workshops that bring together researchers from different fields like automotive engineering, transportation engineering, electrical engineering and computer science.

In this chapter, we first present the motivation and opportunity of implementing VANET-based applications with the advantage of improving level of service provided and we show a possible classification of those applications. At next, we introduce epidemiology-based works in connection with information spreading in order to find an analytical model for information dissemination. We present vehicular mobility models and introduce daily variation of road traffic which both expected to effect performance of information dissemination. We present a survey of state-of-the art works of VANET community related to dissemination strategies and finally we show realistic simulators for VANET.

2-1 Vehicular Applications

Currently used traffic information systems are centralized vehicular applications using technology like Traffic Message Channel (TMC), which provides information about road traffic conditions. However, it is (i) lacking of short delay times (due to the centralized approach) and (ii) averaging information for large geographical areas (due to cost-sensitiveness of detailed sensor networks and limited radio resource) (iii) without the opportunity of providing services for locally interesting and time-critical applications [1]. Moreover, as discussed in [2], the implementation for complete coverage would require new infrastructure which is cost-sensitive as shown in [3]

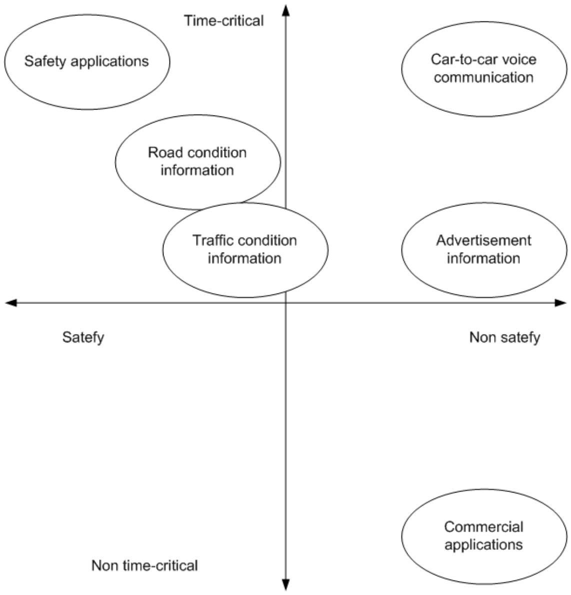


Figure 2-1: Classification of Vehicular Applications based on their time-criticalness and improvement of safety

through a case study. Such systems would for example not meet requirement of an accident avoidance application, because they have large delays and they would require large capacity due to the large geographical area of service. In contrast, VANET-based systems can have short delays and the capacity can be reused more efficiently. Moreover, the structure of VANET can be distributed which improves the level of independence, scalability and stability.

Vehicular application can be classified based on their improvement of safety and time critical nature of the service as shown in Figure 2-1. The figure shows examples of application categories such as safety applications (e.g., avoidance of collision, information of loss of control of vehicle) which can improve road safety significantly (new level of road safety assistance), but is highly time-critical. Comparatively, service about traffic condition information is less time-critical, as discussed in [4], due to the low level of variation of traffic condition in a short time (traffic jams have to build up) and it has a lower impact on improvement of road safety than for example traffic accident assistance applications. This categorization can be extended with expected improvement of safety-comfort level (e.g., example for safety: intersection collision avoidance; for comfort: dissemination of free parking places in large parking lots [5])

More detailed classification of Vehicular Safety Communication-Applications (VSC-A) to seven safety applications is shown in [6]:

- Emergency Electronic Brake Lights: warns of sudden breaking of vehicles in the forward path.

- Forward Collision Warning: warns of possible collision to fortrunning vehicle.
- Blind Spot Warning/Lane change Warning: in case of lack of visual contact due to blind stop it warns during a dangerous lane-change.
- Intersection Movement Assist: warns when it is not safe to enter intersection (e.g., information of traffic light applies to the current vehicle, information about a vehicle which might neglect the traffic light and will probably enter to the intersection).
- Do Not Pass Warning: warns when oncoming vehicle poses collision threat.
- Control Loss Warning: Self-generated warning when vehicle loses control.

We find it important to note the importance of understanding requirements of vehicular applications. Different applications might have significantly different need on properties of information spreading (e.g., delay, distance). Therefore different dissemination strategies (including technology employed) need to be employed for various vehicular applications. We aim to find an analytical model to study the effect of these diverse dissemination strategies on information spreading.

2-2 Epidemics

Researches of epidemiology, physics, biology study dissemination from the point of view of infection spreading. Mathematical models in epidemiology are based on the Susceptible, Infectious, Removed (SIR) model. This model divides the population into three compartments as follows:

- Susceptible (S): individuals without immunity in case of meeting any infectious agent.
- Infectious (I): infected individuals who can transmit the infection to susceptible agent at contact.
- Removed (R): immune individuals who do not affect the dynamics of infection transmission.

The total population is N, and $N = S + I + R$ as introduced in [7]. In case of large enough N, variable S,I and R can be seen as continuous variables and the model can be expressed with the following differential equations [8] [9]:

$$\frac{dS}{dt} = -\lambda * S * I, \tag{2-1}$$

$$\frac{dI}{dt} = \lambda * S * I - \beta * I, \tag{2-2}$$

where the transmission rate λ and β stand for the recovery rate.

Equations 2-1 and 2-2 do not account for distance which is a general case for epidemiology. Advanced partial differential epidemic model [8] show spatial structure of S, I and R, but it assumes random movement of vehicles and suggests to use discrete space models (path models) Those models account for homogeneous mixing, individuals grouped to cells where nodes fully-connected which is not the case for vehicular ad-hoc networks.

Studies of critical epidemics in lattice site structure introduce critical rate o_c that an infinite epidemic will occur, based on infection to nearest neighbors rate o. They further show that the problem of infinite epidemics is a percolation problem where infinite infection is not possible for $o < o_c$, but the probability of infinite epidemic is $0 < P(o) < 1$, if $o > o_c$ [10]. Clusters grow only to a finite size in case of finite epidemic [11]. These studies are assuming recovery of individuals and do not account for distance, but presents that epidemic type of spreading shows percolation.

Bond percolation theory in lattice shows that for a given p probability, that connection exist between edges, exist a critical p_c above which the probability of having an infinite open cluster is always one, or if it is below p_c than the probability is zero. For square lattice it is shown in [12] that $p_c = 0,5$. Directed percolation on square is studied in [13] in order to calculate cluster size (percolation probability exponent) with the help of series expansion of the Catalan numbers.

Complex network theory takes into account small-world effect and scale-free property which has been found in nature for several cases as shown in [14]. In [15] it is suggested to study dynamics of epidemic spreading for different topology structures and heterogeneous population (suggested in [16] as well) while accounting for only infectious(I) and susceptible(S) individuals with scale-free degree distribution. Moreover, it summarizes related work on small-word networks that exists a critical threshold under the infection dies out, in contrary to scale-free networks where dissemination is influenced by low spreading rates as well.

2-3 VANET Dissemination Strategies

Most VANET-based systems assume priori knowledge about the underlying road network which is usually interpreted as a weighted graph [1] [17]. Common approach is to divide the roads to sections with different weights, but not certainly with the same length. The weights are given according to a certain property which can be physical like message's traversal delay [1] or stochastic probability based on the distance of vehicles [18].

Vehicles are assumed to be equipped with sensors which are providing data about the status of the vehicle e.g., speed, geographical position, temperature [19] or even sensors to detect bumps, acceleration [20] or honking [21]. This status represents local information about a geographical area at a certain time moment. Distribution of local information needs to be detailed within closer vicinity, and coarser with the increase of distance as proposed in [22] [23]. For example a driver would be interested in the average speed of vehicles way ahead, but the exact speed of a vehicle 100 meters ahead − to be able to avoid a collision.

Based on the type of communication three main categories can be introduced. First, vehicles sending their messages via a cellular system – and/or Road-Side Unit (RSU) – to a central server or to another peer as described in [24] [25]. The disadvantage of such systems is the high cost of construction and maintenance of the infrastructure. Second group can be the group of systems which are not using cellular systems, but another dedicated like Urban Multi-hop Broadcast protocol suggested in [26] or more general communication technologies for VANET (e.g., Wi-Fi) [27] [28] [29] [30]. Last, a hybrid solution, combination of both systems, seems to be the most powerful, but the most complex approach. Such a system is introduced in [31], to enhance radio coverage of the fix infrastructure (e.g., RSU) by cooperation with vehicles to forward messages from and to the infrastructure. Another approach is to employ static-nodes at intersections to improve the performance of vehicle-based (ad-hoc) information dissemination as shown in [32]. These hybrid systems are taking advantages of ad-hoc- and infrastructure-based communication approaches in order to optimize information spreading.

Fundamental idea of information dissemination for VANET is to have periodic broadcast messages, as presented in [1], and have event driven messages for causes of emergency situations. Vehicles are most of the times sending messages about their current status (velocity, heading) and/or knowledge about the network performance (e.g., delay of certain links, density of cars at a road section). Data from multiple inputs are being processed and a new message calculated and transmitted if the routing protocol requests it. Aforementioned information should be aggregated to fulfill scalability requirements.

Flooding is not scalable though it consumes high amount of energy, bandwidth and memory space while being inefficient [33]. Therefore techniques to reduce network load are required. The main goal is to provide less information with higher distance to keep the system scalable as shown in [34] [35]. Atomic information (e.g., velocity, degree) [36] is being aggregated with information from another nodes [1] or about road sections [17] to have aggregated messages. The message has to be aggregated with new information of the current node before another broadcast takes place.

For aggregation, geographically-hierarchical approaches are being introduced. Level of hierarchy can, for example, be represented in different resolutions of the road network between landmarks as suggested in [23]. Another approach is to form grids of different sizes based on various size of geographical areas as shown in [37]. In both cases the amount of information about a faraway geographical area is reduced, therefore data needed to be maintained as well. Dynamic store and forward approaches are being used to maximize the probability that two message are going to be present in the same time at current node and it will have a chance to aggregate them into a single message and transmitted.

We found an evolution from flooding and directed broadcast transmission to the target area [24] [26] [38] [39] to a distributed (Content Adressable Network (CAN)-based), subscriber/publisher approach [29]. The nodes are going to subscribe to certain groups of information (for example information about certain road sections on the route ahead) and get push messages at a distributed fashion (dissemination is distributed as well as the location of the stored data). Vehicles are sending (pushing into the network) information about their observations, abrupt events. However, as suggested in [40] dissemination strategy might not need to be a single strategy, but a combination of strategies due to the different scenarios of information dissemination for vehicular applications.

Data is preferably stored in a distributed fashion like the CAN which have a d dimensional key space used to give the address of a certain area based on its geographical position. It gives the possibility to reach nodes, which are responsible to store data about a certain area, without knowing their address. For example in case of searching for a certain geographical area, the process will use the geographical position of the area of interest to generate its address. Improvements are being proposed in [25] that (i) new nodes are responsible to store information about areas which are being queried more frequent (load balancing, scalability improvement) or (ii) caching the route of search, because of the high chance that an information about an adjacent road section can be easily reached from the previous route's address or (iii) preprocessing data about a group of road sections which are following each other (for example a highway route).

2-4 Mobility

Studies of Vehicular Ad-Hoc Network (VANET) are mainly based on results of simulations, because real-world testbeds are impracticable while simulations are fast, repeatable and provide opportunity to compare different system approaches. In this section, we present relevant mobility models for vehicular simulations.

Movement of vehicles depends on a wide range of factors such as road structure, traffic rules (e.g., speed limit, traffic lights, on-way roads), traffic conditions, weather conditions, human behavior. VANET community researches are focusing on modeling these factors together in an efficient way. As a result, it is expected that realistic simulations based on vehicular mobility models are mimicking movement patterns well. Different mobility models need to be used for different scenarios of applications.

2-4-1 Random Models

Random models were favored over other model, because of their implementational simplicity and their stochastic properties which allows conducting analytical studies easily. Most popular random models are Random Waypoint Model (RWP) [41] and Random Walk Model (RWM) [42]. In RWP, nodes are randomly choosing the position of their next destination while nodes following the RWM are randomly generating a direction and a travel time for their next journey. However, none of these models are modeling real-world mobility, but they are favored for their simplicity and not for their accuracy. For example in [43] it is shown through analysis that for RWP, increasing number of nodes results in a higher chance of connectivity and a lower minimum transmission range.

Simulation is improved if the movement of vehicles are restricted to a real road network similar structures. Nodes moving according to the Freeway and Manhattan Model are restricted to travel on bidirectional multi-lane freeways, on the edges of a lattice respectively. In case of the Manhattan Model the vehicles are using stochastic turning function to find direction of next movement. Härri et al. classify vehicular mobility models in four classes [44]:

- Synthetic models: Mathematical modeling of movements based on rules and interactions among vehicles which often results in too complex or unrealizable models as presented in [44].
- Survey-based models: Based on statistical results of large scale surveys, mobility models are reproducing random or deterministic behaviors of real traffic. That kind of data is often used for validation purposes as shown in [45] as traces of attendees of Infocom 2005 were used for validation. In [46] is proposed to do coordinated experiments in order to study dangerous situations in an efficient manner (vehicles heading to an intersection at a way that they might collide).
- Trace-based models: Such models are generalizing movement traces to generic movement patterns. An example of an analytical model of such is based on employing public transportation and is shown in [47].
- Traffic Simulator-based models: Some realistic and validated (through results of surveys and traces) vehicular simulators are modeling traffic well enough to couple them with network simulators and extract vehicular behaviors accurate enough.

2-4-2 Towards More Realistic Models

The random models are not enough detailed to study higher level of interactions like traffic jams. They are lacking physical interactions among vehicles and interaction to environment like the flow-density relationship. Studies show that the flow of vehicles can be classified in three groups based on the flow-density relationship: (i) free-flow with rare interaction among vehicles, (ii) coexistence phase in which the flow is reaching its maximum at a critical density and (iii) congested traffic situation with traffic jams and lower level of flow than at the critical density was [48] [49] [50]. Flow based models can be applied in different level of detail as discussed [51] [52]. More realistic models can be classified as follows:

- microscopic: Entity based model where all parameters of vehicles are considered.
- macroscopic: Inspired by fluid theory in order to find macroscopic meanings of mobility.
- mesoscopic: Intermediate level between the models above to take advantage of a scalable, but detailed enough model.

Microscopic models are highly detailed while macroscopic models are grouping vehicles with similar parameters. The level of detailedness is a tradeoff though microscopic models are computationally complex compared to macroscopic models.

Realistic movement rules are being modeled by taking into account models with specific functions like the Car Following Model (CFM) where the vehicles need to keep a safety distance through deceleration or the Intelligent Driver Model (IDM) which aims to hold the desired speed of traveling without reaching a limit of gap between vehicles. Changes of parameters of roads are also modeled. For example accounting for traffic lights are considered in intersection management models.

Survey based models can further be grouped to path and trip models. Trip models are based on movement of vehicle(s) between points of their interest while path models are based on individual paths of vehicles traveling between. Mobility traces are used to calibrate survey- and traffic simulator-based models.

Measurements by sensors on roads and surveys are showing a great variance of the number of vehicles by the hour of day studied [53] [54] which has a high impact on traffic conditions and might have effect on services requested at specific ours of a day (for example less detailed need for traffic condition information by night due to low traffic condition, but constant request for road condition information like icy road section).

2-5 VANET Simulators

Vehicular simulators are tools to generate mobility traces from vehicular mobility models. These traces are than used for simulating existing or new protocols. They provide the opportunity to developers to do fast, well-configurable, cheap and repeatable simulation runs in order to evaluate results of new protocols or even whole systems. Simulations are not accounting for all the parameters of reality, but for selected ones to keep simulations feasible while keeping them accurate enough.

Simulators need to combine vehicle networking and vehicular traffic simulation [55]. Either they both incorporate to a single simulator application like presented in [56] or the traffic simulator outputs traces for a networking simulator which than provides results of traffic and communication performance (e.g., average speed) and networking characteristics (e.g., inter-contact time, minimum transmission range [57], mean time to loss information [58]).

Simulator of Urban MObility (SUMO) is a traffic simulator capable of importing user defined road network structure and simulate single vehicle's movement traveling on the road network (micro mobility).

The Network Simulator (ns) version two (ns-2) is a network simulator aiming to simulate discrete events. It supports TCP, routing and multicast protocols as well. It is used for evaluation in [55] [59].

VanetMobiSim is a vehicular mobility simulator supporting macro and micro mobility models [60] [61]. Macro-mobility model is taking into account the road structure (single- or multi-lane), topology, characterization (speed limit) and the movement pattern provided either by the trip or path generator module. Parameters for car's speed and acceleration modeling are individually available for the micro-mobility model as well as three various level of details for computing vehicle's speed based on (i) deterministic process, (ii) single-lane scenario with interaction among vehicles or (iii) multi-lane scenario and function of nearby vehicles. Traces of VanetMobiSim are validated by TSIS-CORSIM.

Chapter 3

Lattice-based Dissemination: Analytical Modeling

In this chapter, two theoretical cases of information spreading are being defined and studied from an analytical point of view. Therefore, we define a simple realization of information dissemination called Lattice model and introduce the *general* and *first-wave* theoretical cases which we analyze. The two cases are being studied through recursive type of models, upper and lower bound approximations and numerical simulations. Numerical evaluations of the models and approximations are presented as well as a mathematical model approximating results of numerical simulations of the First-wave case.

3-1 Modeling Dissemination

Analytical modeling of spreading processes relies mostly on Susceptible, Infectious, Removed (SIR) models as presented in Section 2-2 and in [7] [8] [9]. SIR models are not accounting for distance, but for time and λ, being the rate of infection and provide the number of individuals being infected and the sum of all individuals. SIR(t, λ) = % of individuals infected. Moreover, SIR models do not provide possibility to take into account the structure of road network and vehicular movement patterns, but it assumes homogeneous mixing of vehicles.

In order to account for distance while allowing analytical tractability, we divided the space with a lattice, which represents a grid-like road structure. All positions are being associated to a cell. The most important variable of a cell is its distance from the center position which is called its distance as shown in Figure 3-2. This

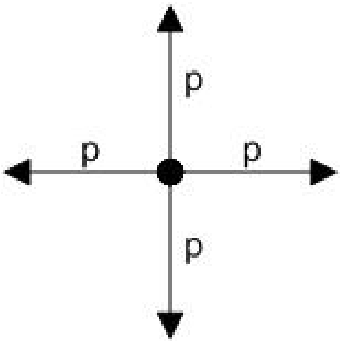


Figure 3-1: Four direction of transition probability p

notation of distance is being used to define the distance from the source position of the information spread. Further, we assume a homogeneous transition probability p to all four directions. Transition probability p abstracts transmission among vehicles at different cells or carrying information between cells. It is ranging as $(0, 1]$ to all four directions independently, see Figure 3-1. We call the above introduced model the Lattice model throughout our work.

The Lattice model is simple enough to study analytically and numerically, but more realistic than if we would consider vehicles following the Random Walk Model or Random Waypoint Model though the road network is not random. It is more like network of parallel and crossing roads.

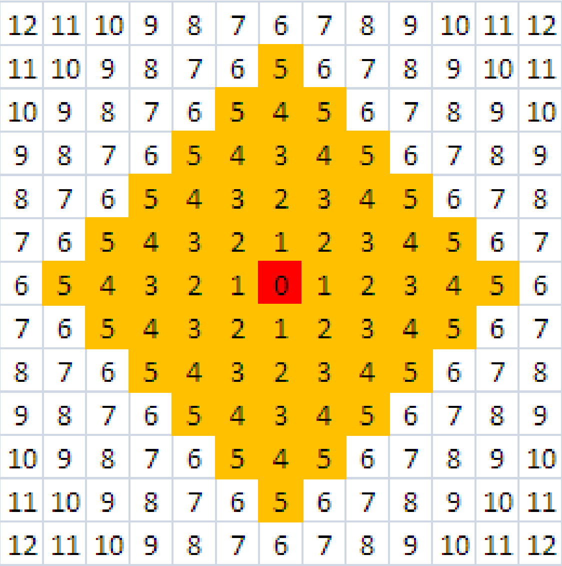
We present two cases we considered and studied as follows:

- General Case: Spreading of an information according to the lattice model. We are interested in the number of nodes aware of the information at a certain distance and time moment, see Figure 3-2. The *general* case only accounts for nodes at a given distance and not for all nodes within the given distance.
- First-wave Case: It is based on the *general* case, but only accounts for nodes at the maximum distance the information could travel for a given time. We study the theoretical boarder of the information dissemination. Therefore, time and distance is numerically equal for the *first-wave* case, see Figure 3-3.

The number of all nodes at a certain distance can be written as

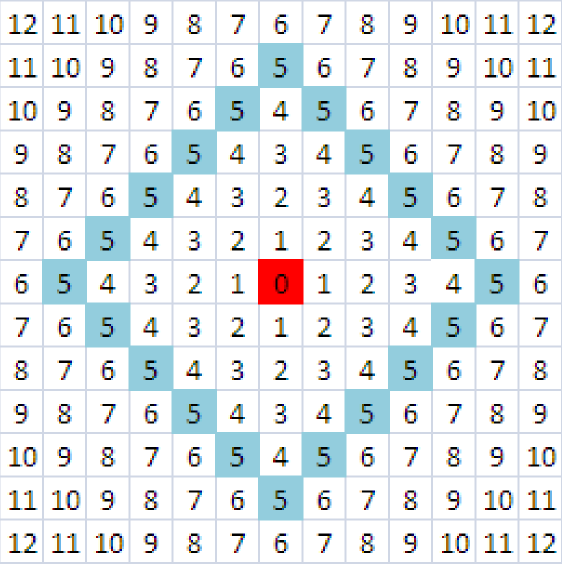
$$N_i = 4 * i, \tag{3-1}$$

where N_i is the number of all nodes at distance i if $i > 0$ and $N_1 = 1$ as shown in Figure 3-3.



12	11	10	9	8	7	6	7	8	9	10	11	12
11	10	9	8	7	6	5	6	7	8	9	10	11
10	9	8	7	6	5	4	5	6	7	8	9	10
9	8	7	6	5	4	3	4	5	6	7	8	9
8	7	6	5	4	3	2	3	4	5	6	7	8
7	6	5	4	3	2	1	2	3	4	5	6	7
6	5	4	3	2	1	0	1	2	3	4	5	6
7	6	5	4	3	2	1	2	3	4	5	6	7
8	7	6	5	4	3	2	3	4	5	6	7	8
9	8	7	6	5	4	3	4	5	6	7	8	9
10	9	8	7	6	5	4	5	6	7	8	9	10
11	10	9	8	7	6	5	6	7	8	9	10	11
12	11	10	9	8	7	6	7	8	9	10	11	12

Figure 3-2: General case with lattice structure: number of all cells aware of information at a certain distance (but the distance equals for the cells studied) depending on the time limit t of spreading and transition probability p. Notation of distance starting with the red colored source cell. Only distances within the orange field studied, because that is the maximum the message could travel within time limit $t = 5$.



12	11	10	9	8	7	6	7	8	9	10	11	12
11	10	9	8	7	6	5	6	7	8	9	10	11
10	9	8	7	6	5	4	5	6	7	8	9	10
9	8	7	6	5	4	3	4	5	6	7	8	9
8	7	6	5	4	3	2	3	4	5	6	7	8
7	6	5	4	3	2	1	2	3	4	5	6	7
6	5	4	3	2	1	0	1	2	3	4	5	6
7	6	5	4	3	2	1	2	3	4	5	6	7
8	7	6	5	4	3	2	3	4	5	6	7	8
9	8	7	6	5	4	3	4	5	6	7	8	9
10	9	8	7	6	5	4	5	6	7	8	9	10
11	10	9	8	7	6	5	6	7	8	9	10	11
12	11	10	9	8	7	6	7	8	9	10	11	12

Figure 3-3: First-wave case with lattice structure: number of all cells aware of information at distance the information could maximum travel in time $t = 5$ (blue colored) and source of information (red colored). Distance and time is numerically equal.

3-2 Recursion Based Models and Approximations

The nature of our model implies the practical use of a recursive model to find the number of nodes being aware of the information. We consider two cases: the *general* case where we are interested in the information awareness of cells at certain distances only and within the maximum transmission distance (limited by upper limit of time). In case of the *first-wave* case, only cells are considered which are at the maximum distance the information could reach within the time limit.

3-2-1 General Case

Nodes of a cell can get the information from only their neighboring cells based on the declaration of the Lattice model, see Section 3-1. Number of neighboring cells is four, two cells at a previous distance and two cells at distance one further. Based on that assumption we introduce the Recursive General model as follows:

$$I_i(t) = I_i(t-1) + [N_i - I_i(t-1)] \left[2p \frac{I_{i-1}(t-1)}{N_{i-1}} + 2p \frac{I_{i+1}(t-1)}{N_{i+1}} \right], \qquad (3\text{-}2)$$

where $I_i(t)$ is the number of nodes aware of the information at distance i, at time moment t. Distance is measured in number of cells from the information source while time is running from the time moment the information spread started. Therefore, both variables are discrete and positive $(i, t > 1)$. Number of all nodes at distance i only is N_i, see Eq. (3-1). Transition probability $p \in \mathfrak{Re}$ and $0 < p =< 1$.

We consider that all nodes can get the information from four neighboring nodes which is not true for at least the four nodes at the largest horizontal and vertical distance (border effect). Therefore, it is an over estimation of the problem which is remarkable for short distances (for high distances it could be neglected if the recursive calculation for short distances are correct). For example in case of $p = 1$ the number of all nodes at distance 10 should be 40, see Eq. (3-1), but based on the General Recursive model we got 1024, see Eq. (3-2). The difference is remarkable and makes the model unfeasible to use.

It can be seen from Eq. (3-2) that a remarkable simplification is possible if we consider the First-wave case of dissemination though we can neglect all influences from nodes at greater distance than the studied one $(I_{i+1}(t-1) = 0)$ and we make the assumption that nodes at equal distance cannot send the information (simple they cannot be aware of it, yet)$(I_i(t-1) = 0)$. Therefore, we continue our study in the next section focusing on the *first-wave* case and use the findings to study the *general* case.

3-2-2 First-wave case

For the *first-wave* case, Eq. (3-2) can be simplified. Number of all cells aware of information at the same distance, but at a time moment earlier ($I_i(t-1)$) is zero ($I_i(t-1)=0$) due to the fact that we account for the *first-wave* case where information cannot be present at an earlier time moment ($t-1$) than t, at the same distance (I_i). This implies

$$[N_i - I_i(t-1)] = N_i \tag{3-3}$$

Moreover, number of all cells aware of information at distances higher than the studied one is

$$I_{i+z} = 0, \tag{3-4}$$

where $z \in Z^+$. Those cells are at distances the message could not travel yet ($i > t$ implies $I_{i+z} = 0$). At last, the simplified recursive equations, based on the simplifications suggested above, follows

$$\widetilde{I}_i(t) = N_i \left[2p \frac{(\widetilde{I}_{i-1}(t-1))}{N_{i-1}} \right], \tag{3-5}$$

where $\widetilde{I}_i(t)$ stands for the number of nodes being aware of information at time t and at distance i, if the transition probability is p for the *first-wave* case.

We further take the assumption:

$$N_i = N_{i-1} \qquad \text{, if} \quad N_{i->\infty} \tag{3-6}$$

and the initial value follows:

$$\widetilde{I}_1 = 4p \tag{3-7}$$

Recursive model for the *first-wave* case based on the simplifications and assumptions is the following:

$$\widetilde{I}_i = \widetilde{I}_{i-1} 2p \tag{3-8}$$

Finally, the Recursive First-wave model:

$$\widetilde{I}_i = 4p(2p)^{i-1}, \tag{3-9}$$

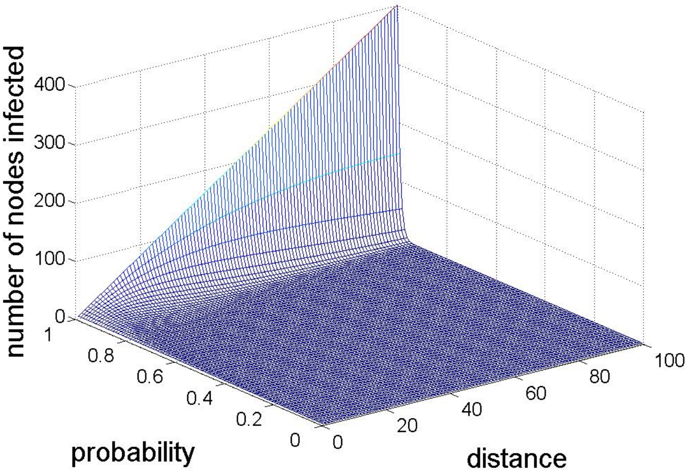


Figure 3-4: Compensated Recursive model for First-wave case. Showing the number of nodes aware of information for a certain transition probability and distance variable if time limit $t = 100$. Spreading is only feasible for very high p transition probabilities.

which tells the number of cells aware of information at distance i if time limit is t, for the *first-wave* case with transition probability p.

The model is based on the Recursive General model which we showed is an overestimation of the problem which applies to the Recursive First-wave model as well. The border effect, we did not consider, leads to a remarkable overestimation. Therefore, we show a possible compensation of the overestimation in the next section.

3-2-3 Compensated Recursive model for First-wave case

The Compensated Recursive model is based on the Recursive First-wave model which is accounting for only the cells at the boarder of information dissemination (*first-wave* case). We propose to compensate for the fact that nodes at distance i cannot infect $2i$ number of nodes which are at distance $i + 1$, but just 4 more nodes as a sum $(4(i + 1) = 4i + 4)$:

$$\widetilde{I}_i = 4p(2p)^{i-1}\frac{i}{2^{i-1}} \tag{3-10}$$

The Compensated Recursive model (above) accounts for the fact that the number of nodes are not increasing by the factor of two to power, but linearly with four times the increase of distance.

Numerical evaluation showed that this model is an under estimation of the problem. For example in case of $p = 1$ the number of all nodes at distance 10 should be 40 as

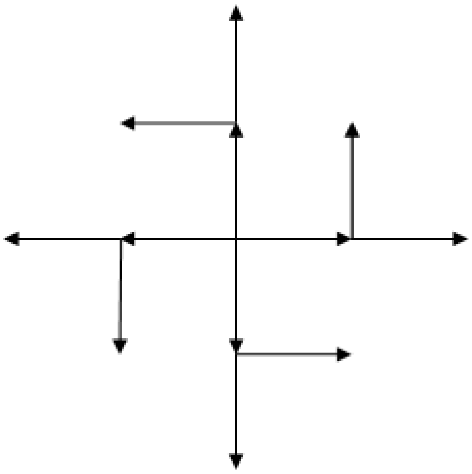


Figure 3-5: Lower bound approximation of First-wave case: cells can get messages from exactly one cell at a previous distance.

shown in Eq. (3-1), but based on the Compensated Recursive model we calculated 20, see Eq. (3-10). The difference is remarkable and makes the model unfeasible to use.

3-2-4 Upper/Lower Bound Approximation

We present an upper and lower bound approximation of the problem of information dissemination for the *first-wave* case.

Lower

A simplified version of the simple model with the assumption that all nodes can get information from exactly one node at a previous distance as shown in Figure 3-5. We found, based on theoretical study, that the expected value of $\widetilde{I}_i$ follows:

$$E(\widetilde{I}_i) = p^i 4i \tag{3-11}$$

Upper

An extension of the Lattice model with the assumption that all nodes can get information from exactly two different nodes at a previous distance as shown in Figure 3-7. The recursive model follows:

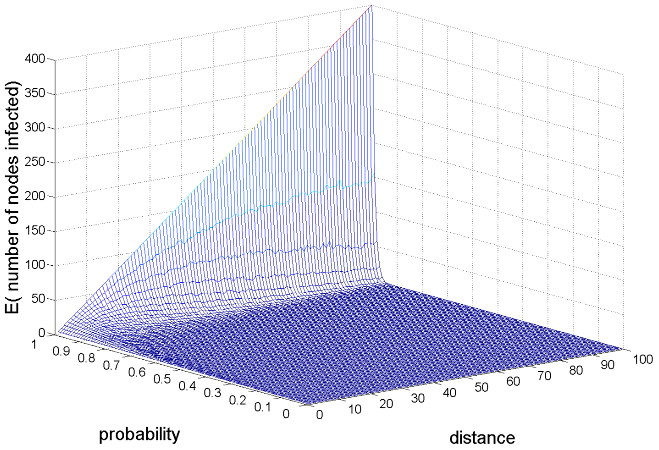


Figure 3-6: Numerical simulation of Lower bound approximation for First-wave case: showing the number of nodes aware of information for a certain transition probability and distance variable if time limit $t = 100$. Results are averaged over 10000 simulation roundsSpreading is only feasible for very high p transition probabilities.

$$\widetilde{I}_i(t) = \frac{I(i-1)}{4(i-1)} 4i2p \tag{3-12}$$

Expected value of $\widetilde{I}_i$ follows:

$$E(\widetilde{I}_i) = p(2p)^{i-1}4i \tag{3-13}$$

3-2-5 Comparison of Studied Models and Approximations

Studied models and approximations of Section 3-2 can be placed in two groups:

Upper approximation

General Recursive model and *upper bound* model is an upper approximate of the *first-wave* case. The General Recursive model is in linear dependency of the *upper bound* approximation based on distance (multiplication with d).

Lower approximation

Compensated model and *lower bound* model are a lower approximation of information dissemination of the *first-wave* case. Calculation based on those models show a lower value for the number of nodes aware of information than theoretical calculation or numerical simulation does. The two model equals.

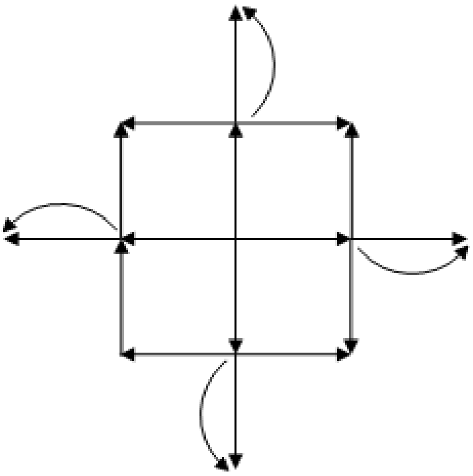


Figure 3-7: Upper bound approximation of First-wave case: cells can get messages from exactly two cells at a previous distance.

3-3 Numerical Evaluations of Lattice-based Models

We developed and implemented our proposed analytical model, the Lattice model, on Matlab by modeling the cells of the simulation's area as positions of a matrix and defined the initial values of cells as zero except for the center position (horizontal and vertical center) of matrix which models the single information source. Cells aware of the information are allowed to spread the information independently to all 4 directions (horizontal and vertical steps) with transition probability p at every time moment. We have used the above introduced numerical simulation model for modeling information spreading for both, the General and First-wave case.

Based on numerical simulations, we found the number of nodes being aware of information at a certain transition probability-distance pair, see Figure 3-8 and Figure 3-10. Figure 3-8 shows the number of nodes aware of the information at distance d if transition probability is p for the *general* case while Figure 3-10 shows the same representation of data for the *first-wave* case. Transition probability $p \in \Re$ and has a range of $p = [0.1, 1]$, time limit $t = 100$ and maximum of distance is $d =$ 100 cells to all four direction. The results are averaged over 10000 simulation rounds at which level the statistical variation is low enough while performing simulations is still feasible (above that level the simulations took significantly more time to run). Matlab is being used to do the simulations. Division of space to cells is being represented by matrix positions and the spreading starts from the center of matrix/space.

We find it important to note that in both cases at $p = 1$ the number of nodes being aware of information is increasing with the increase of distance linearly as we

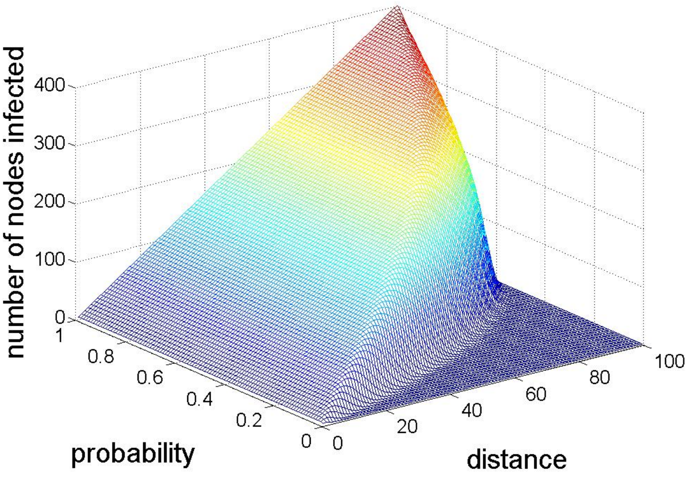


Figure 3-8: General case with time limit $t = 100$. Shows the number of nodes aware of information for a certain transition probability and distance variable. Results are averaged over 10000 simulation rounds. Spreading is not feasible to high distances in combination with low transition probabilities .

introduced earlier in Eq. (3-1). In case of low transition probability p the information spread is not going to reach greater distances, but we found that for higher transition probability p the reachable distance within the same time limit is always going to be greater. In order to increase that distance we need either to increase the transition probability p or to increase the time limit as shown in Figure 3-9, where time limit is twice compared to Figure 3-8.

We observe differences between results of the two models. At first, the distance and time is numerically equal for the *first-wave* case, because of the declaration, see Section 3-1. Moreover, at low transition probability p the spreading is not feasible compared to the *general* model. We would like to note that we found a phase transition around the value of transition probability $p = 0.5$ above which the result shows similar behavior to the *general* model and finally at $p = 1$ the two results are matching, see Figure 3-8 and Figure 3-10. We observe also from the figures that at both cases the number of nodes aware of information for a certain p increases with the increase of distance up to a limiting value of maximum nodes (limited by the numerous rounds of transition probability) and starts to decrease afterwards, except at case $p = 1$, where the increase is continuously linear, see Figure 3-8 and Eq. (3-1) (maximum number of cells at certain distance will be equal to the cells aware of information if $p = 1$).

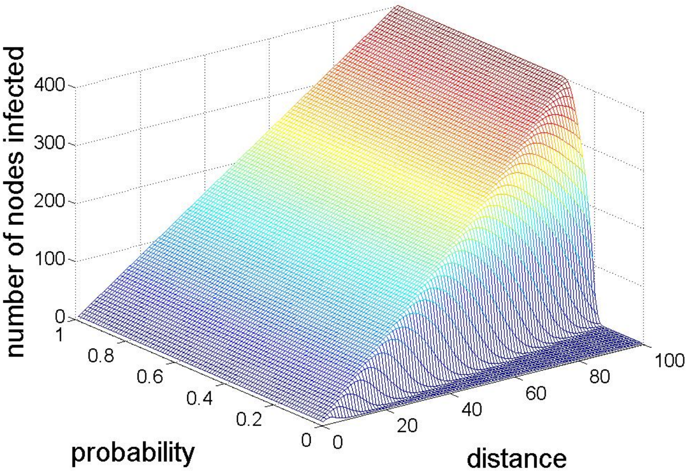


Figure 3-9: General case for time limit $t = 200$. Shows the number of nodes aware of information for a certain transition probability and distance variable. Results are averaged over 10000 simulation rounds. Result of extended time limit (compared to Figure 3-8) shows that spreading is more efficient and feasible for low transition probabilities and high distances.

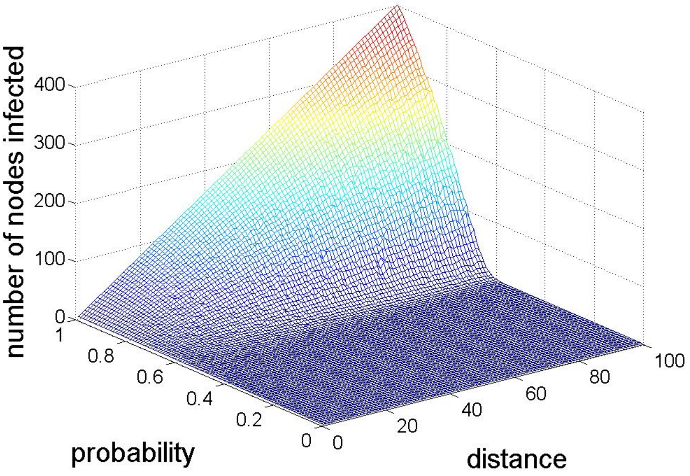


Figure 3-10: First-wave case with time limit $t = 100$. Shows the number of nodes aware of information for a certain transition probability and distance variable accounting for the First-wave case. Results are averaged over 10000 simulation rounds. Spreading is not feasible for low transition probabilities. Phase transition at found at $p = 0.5$

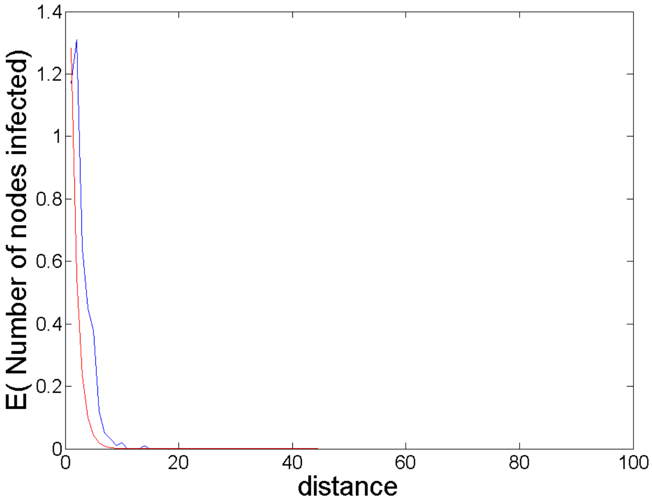


Figure 3-11: Model fitting example for $p = 0.3$: red line is averaged result over 10000 simulation rounds accounting for the First wave case with transition probability $p = 0.3$. $E(\widetilde{I}_{i,0.3}) = 3e^{-0.85i}$. Blue line is calculated based on the Simple mathematical model shown in (3-14).

3-4 Model fitting

3-4-1 Simple

Here, we try to fit our numerical (evaluated) results. Therefore, we define a mathematical equation to approximate the numerical results found in Section 3-3 for the *first-wave* case:

$$E(\widetilde{I}_i) = aie^{-bi} \tag{3-14}$$

where $a, b \in \mathfrak{Re}$ and i is the distance from the source position of information.

We found that the matching with the numerical simulation's result, see Figure 3-11, and calculated values, based on this mathematical model, is not satisfactory for probabilities ranging from $p \approx 0.45 - 0.50$ above. We have introduced these p values as values of the phase transition of the *first-wave* case. Unfortunately, we found that the formula is not able to take into account the initial value at distance one (which is simply $4p$) which leads to the inaccuracy.

3-4-2 Extended

We found an extension to tackle the problem introduced above and we present it for all transition probabilities $p >= 0.58$ as the following:

$$E(\widetilde{I}_{i,p}) = 4p - ae^{-b} + aie^{-bi} \tag{3-15}$$

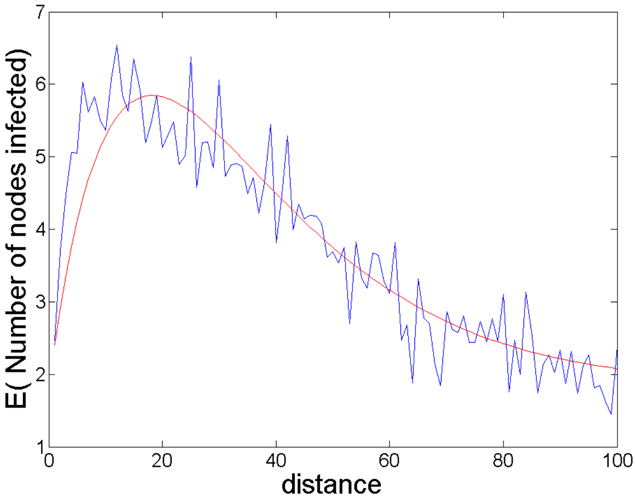


Figure 3-12: Model fitting example for $p = 0.6$: red line is averaged result over 10000 simulation rounds accounting for the First wave case with transition probability $p = 0.6$. $E(\widetilde{I}_{i,0.6}) = 4 \cdot 0.6 - 0.6e^{-0.055} + 0.6ie^{-0.055i}$. Blue line is calculated based on the Extended mathematical model shown in (3-15).

The extension realizes the matching of the initial value, $E(\widetilde{I}_{1,p}) = 4p$.

The reason of the requirement of $p >= 0.58$ is that we found that this is the lowest value of p which satisfies:

$$\lim_{i \to 100} E(\widetilde{I}_i) > 0 \qquad \text{is true if,} \quad p >= 0.58 \tag{3-16}$$

The extended formula works well for transition probabilities higher than 0.58 as shown for two cases in Figure 3-12 and Figure 3-13, but cannot be applied to probabilities lower than that value - where $\lim_{i \to 100} E(\widetilde{I}_i) = 0$ stands true. At probabilities lower than the phase transition we propose to employ the simple model.

In order to evaluate the correctness of the fit we present in Table 3-1 the Root-Mean-Square Error (RMSE) values for the three examples studied ($p = 0.3, p = 0.6, p = 0.75$). RMSE is an absolute measure of fit and it can be interpreted as the standard deviation of the unexplained variance. It is in the same unit value as the response variable (number of nodes) and lower values of RMSE indicate a better fit.

3-4-3 Conclusion

In this chapter, we found two mathematical models which are approximating the number of nodes aware of the information based on time and distance for the *first-wave* case. The simple model can be used for transition probabilities lower than

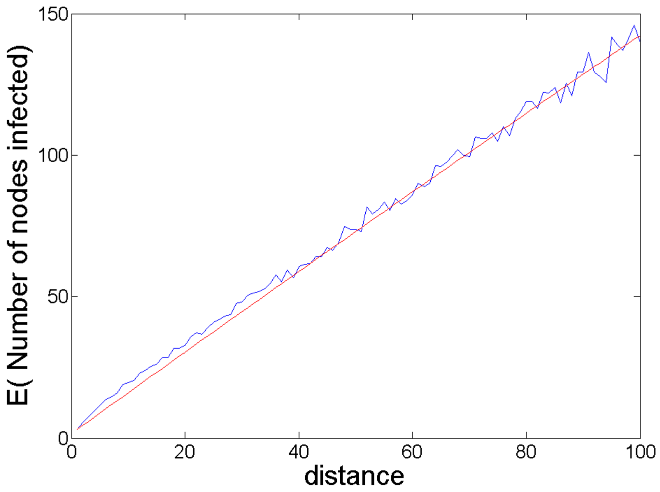


Figure 3-13: Model fitting example for $p = 0.75$: red line is averaged result over 10000 simulation rounds accounting for the First wave case with transition probability $p = 0.75$. $E(\widetilde{I}_{i,0.75}) = 4 \cdot 0.75 - 1.45e^{-1*0.0003} + 1.45ie^{-0.0003i}$. Blue line is calculated based on the Extended mathematical model shown in (3-15).

Table 3-1: RMSE of numerical simulations of the simple and extended mathematical model we proposed in Section 3-4

Transition probability p	RMSE value
0.3	0.100
0.6	0.521
0.75	3.016

phase transition while the extended model approximates well for high transition probabilities by taking into account the initial value correctly.

Chapter 4

From Analytical to Realistic Simulations

In Chapter 3, we introduced a simple analytical model, we used to analyze information spreading on lattice. As a first step, simulations of this chapter are aiming to validate a model equivalent to the simple analytical model, but implemented on the simulator environment. We call that simulator-based model, the *validation* model and we validate its equivalence to the analytical model. We use simulation results of the *validation* model to compare it to the lattice-based analytical model and to do simulations accounting for more realistic mobility models. Realistic simulations are based on real-world road network structure while vehicular mobility is either interaction-free (no dependency on other vehicles) or realistic interaction-based (vehicles movement depend on other vehicles in the proximity). We evaluate simulation results and summarize effects of realistic mimicking of vehicle's movement.

In this chapter, we first present simulation requirements and the Generic Mobility Simulation Framework (GMSF), a realistic vehicular simulator. In order to support requirements of single source epidemic information dissemination we show the extensions we implemented in GMSF and present the realistic road structure scenarios studied. We show steps of simulations and evaluate simulations' results for various combinations of mobility models and scenarios applied.

4-1 GMSF: Mobility Simulator and Extensions

4-1-1 Simulation Requirements

In order to simulate single source information spreading the simulator application needs to be capable of simulating on the microscopic level the movement of vehicles as independent entities with specific parameters e.g., position, communication range, speed, information awareness. Connectivity have to be based on distance between vehicle's position and the communication channel has to be bi-directional and single link to support a fairly realistic car-to-car communication (we are not accounting for realistic communication models though they are out of the scope of this work).

Lattice-based simulations require the Manhattan Mobility (vehicle movement model) and a lattice road structure while realistic simulations require a realistic representation of road structure and realistic mobility models based on interaction among vehicles (e.g., Intelligent Driver Model employing Car Following Model, Intersection Management model). We further require the option of an intermediate step accounting for real-world road structure and interaction-free, stochastic function based movement model (e.g., stochastic turn).

Function of single source information dissemination requires distance notation from starting position of information. A single node is responsible to start information spreading by transferring the information to another vehicles within its radio coverage. At next steps, all vehicles aware of information should relay the message to vehicles within their transmission range. Moreover, the simulator application has to be capable of exporting data about the nodes aware of information in order to post-process results. Finally, GUI of the simulator should provide visual information for better understanding the results (e.g., visualization of infected nodes to study level of clustering).

4-1-2 Generic Mobility Simulation Framework

The GMSF [54] is a generic and expandable vehicular mobility simulator for various mobility models. Therefore, new functionalities and services can be easily added (the source code is available in Java).

GMSF is a tool for evaluating new mobility models compared to well-accepted classical models. GMSF employs GIS-based models relying on maps from the Geographic Information System (GIS) [62] and accounting for realistic microscopic behaviors (e.g., Car Following Model and Intersection Management) [51] and classical mobility models (e.g., Random Waypoint, Manhattan), see Section 2-4 for details of classical mobility models.

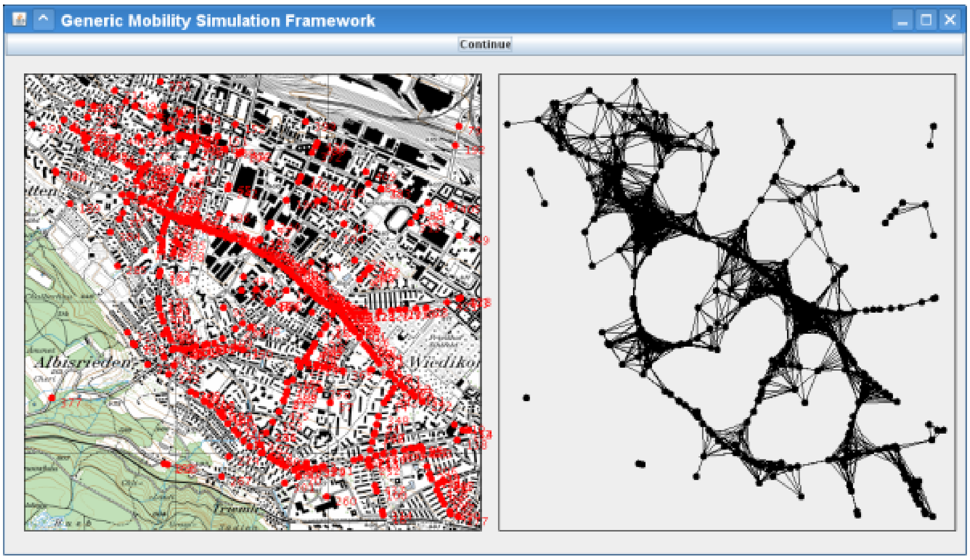


Figure 4-1: Screenshot of GMSF with the GIS-based Urban scenario: Position of nodes on the map (left side) and the corresponding network graph (right side) [54].

Mobility traces generated with GMSF can have various output formats e.g. the mobility trace format of ns-2 [63] or Qualnet [64] network simulators. In addition, traces can be generated in a simulator independent XML-based trace format.

However, the application does not support epidemic type of information dissemination and other requirements we required in Section 4-1-1, but provides visualization of the spreading process, see Figure 4-1 and an easily expendable structure. Therefore, we present at the next chapter the extensions we implemented in GMSF to support our simulations.

4-1-3 Extensions

In an effort to simulate epidemic type single source information spreading we have extended the GMSF application's functionality with the following functions, services:

Single source epidemic dissemination

Only a single node is being aware of the information and deployed on the road network to start information spreading. The position is the center position for lattice-based road structure and random for real-world road based structures in order to simulate the random nature of the spatial source of information.

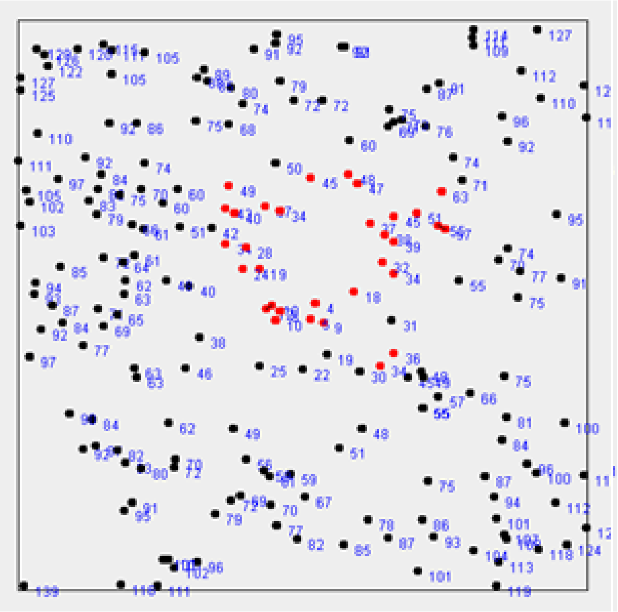


Figure 4-2: First part of screenshot sequence of GMSF accounting for lattice based road structure and Manhattan Mobility. Coloring of the infected nodes (red) is an extension we implemented in GMSF. Nodes colored blue are not aware of the information.

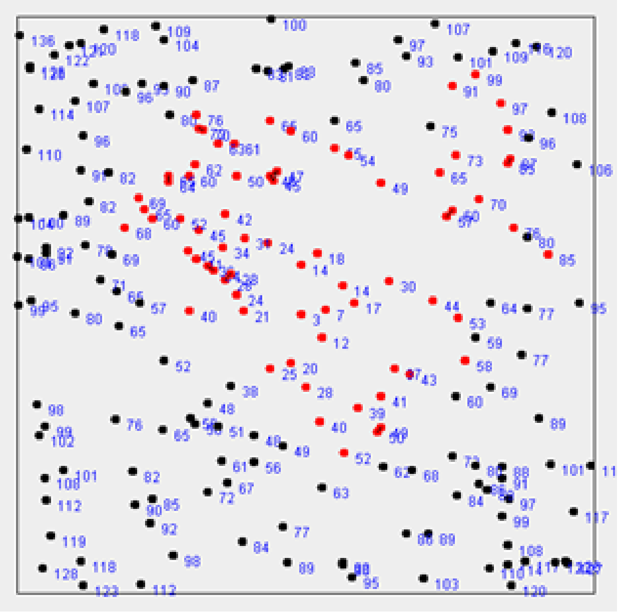


Figure 4-3: Second part of screenshot sequence of GMSF shows how the position and information awareness of nodes changed with time. It need to be compared to Figure 4-2. Coloring of the infected nodes (red) and distance notation shown are an extension we implemented in GMSF. Nodes colored blue are not aware of the information.

Distance from position of information source

The previously mentioned source position is used together with the current position of a node to calculate its Euclidean distance which is than its distance from the source of spreading.

GUI changes

Visualization is being improved and extended. Nodes aware of information is being colored and the distance from the starting point of information spread can be shown on the GUI interface.

Exporting data

The position of the nodes which are aware of information together with a time stamp is written to a file. This data is being post-processed in Matlab.

Multiple simulation runs

The randomized process needs multiple runs to provide statistically satisfying data.

4-1-4 Overview of Simulations

Validation

We introduce the *validation* model which is equivalent to the analytical model introduced in Chapter 3, but implemented in the GMSF environment. In order to support the validation, we implement a model equivalent to the Lattice model by using lattice road structure and the built-in function of the Manhattan Mobility (vehicles are traveling on a grid-based road structure and using stochastic turn function at every intersection reached to decide for next direction of movement) and we set the speed of individual vehicles to equal in order to avoid interaction among them. We call the models and their settings we used in this simulation together, the Validation model.

Realistic simulation

This simulation is accounting for real-word road network structure and interaction-free (among vehicles), stochastic-based mobility model (e.g., stochastic turn function at intersections). We call this simulation realistic for the reason of being based on real-world road structure although realistic vehicular mobility is not being applied. We considered three scenarios of real-world road networks structures e.g., *City*, *Rural* and *Urban*.

Extended realistic simulation

Previously introduced three, real-world road structures are being employed in this simulation in combination with realistic vehicular mobility models provided by the GMSF application (e.g., Car Following Model: deceleration to avoid collision with fortrunning vehicle; traffic lights at intersections controlled by the Intersection Management model). These models applied together are representing the highest level of realism in our simulations.

4-2 Validation Model

We have implemented an equivalent model to the analytical model on the GMSF simulator and we are calling it the *validation* model. For this simulation, we have used the built-in function of GMSF of the Manhattan Mobility model (movement model) and we have chosen the central position for the information source node. The area of simulation was set to 200 by 200 cells and no additional movement rules (e.g., intersection management) were applied. Moreover, we set the transmission range smaller than the length between two intersections in order to keep the spreading based on the movement of vehicles.

Figure 4-4 shows the number of nodes being aware of information at distances of [1,100] cells (collected from nodes at all 4 directions) for 1000 nodes. Time variable is ranging from [0,2000] seconds. It can be seen from the figure that there is a minimum time needed for information spreading to reach certain distances from the source and that the distribution of nodes between different distances is linearly increasing with raise of distance which is shown in (3-1). Moreover, the homogeneous deployment of nodes and their random movement implies that the number of nodes at certain distances are constant if the time constant is large enough. The variation seen on the figure is based on the statistical variation of the result through it is mean value of 100 simulation runs. We have calculated the sum of all nodes aware of information at two different time moment in order to proof the results of the simulation. Table 4-1 shows the number of all nodes being aware of the information at time movement 1000 and 2000 which is 513 and 510 respectively. This number is half the number of all nodes, because we only take into account nodes up till the distance of 100 cells while half of all the nodes are outside of that area, see Figure 3-8.

For the same reason we can see the maximum of nodes within certain distances cannot be larger than the number of all nodes, which is at Figure 4-6 is 1000.

Table 4-1: Simulation results of the Validation model shows half of the nodes simulated are at the are studied and for long time limits the information awareness if complete, all vehicles aware of the information

Time moment	Number of nodes aware	Sum of all nodes
1000	513	1000
2000	510	1000

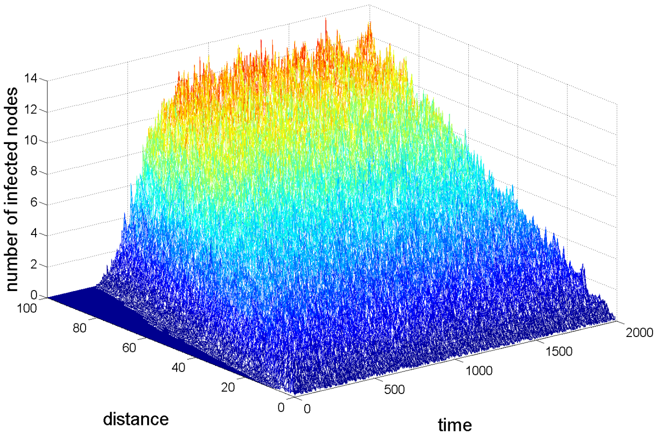


Figure 4-4: Simulation result of the Validation model based on lattice road structure and Manhattan Mobility shows the number of nodes aware of information for certain distance and time variables (not accumulated for distance). Number of simulated nodes is $n = 1000$ and results are averaged over 100 simulation rounds. Spreading for high distances have a minimum requirement on time limit.

Therefore the maximum value of nodes aware of the information within the distance of 100 cells is 500.

Figure 4-5 shows the connection between the number of nodes being infected at different number of all nodes participating at the simulation (ranging from 100-1000) and at different distances ranging from 1 to 100 cells. This representation of data is the same as it can be found at Figure 3-8, although we have used density (we have used the sum of all participating nodes numerically, because the area of simulation is closed and the total number of nodes is constant) instead of transition probability. For our case density is in a linear connection with the transition probability, because the deployment of nodes are homogeneous and the transfer of information is based on the meeting frequency of nodes which is in linear connection with the density of vehicles (for twice as many nodes we would have twice the chance to transfer our information though the chance of having a node at a neighboring cell is the double as well).

It can be seen in both figures that number of nodes aware of information is linearly increasing with increase of distance if the density (sum of all nodes) are large enough.

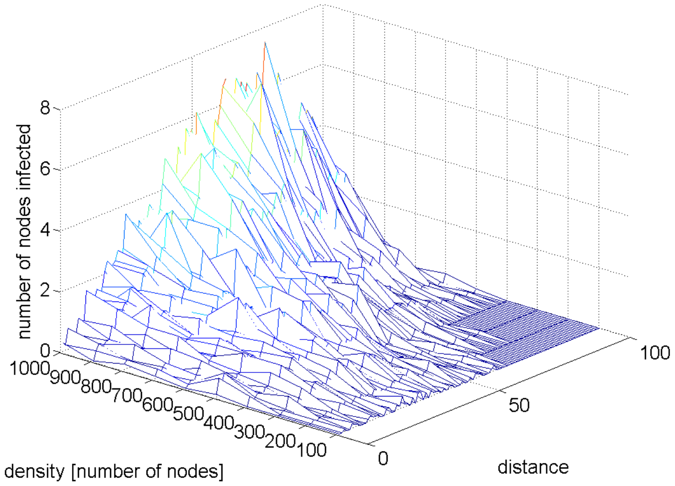


Figure 4-5: Simulation result of the Validation model based on lattice road structure and Manhattan Mobility shows the number of nodes aware of information for certain density and distance variables (not accumulated for distance). Time limit is $t = 250$ and results are averaged over 100 simulation rounds. Spreading is not feasible to high distances in combination with low number of vehicles, low densities respectively.

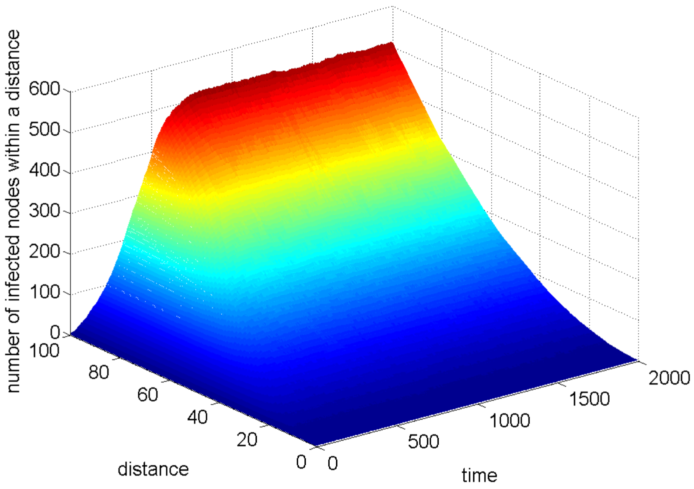


Figure 4-6: Simulation result of the Validation model based on lattice road structure and Manhattan Mobility showing the accumulated number of nodes aware of information at a certain distance with time limit t. Number of simulated nodes $n = 1000$ and results are averaged over 100 simulation rounds.

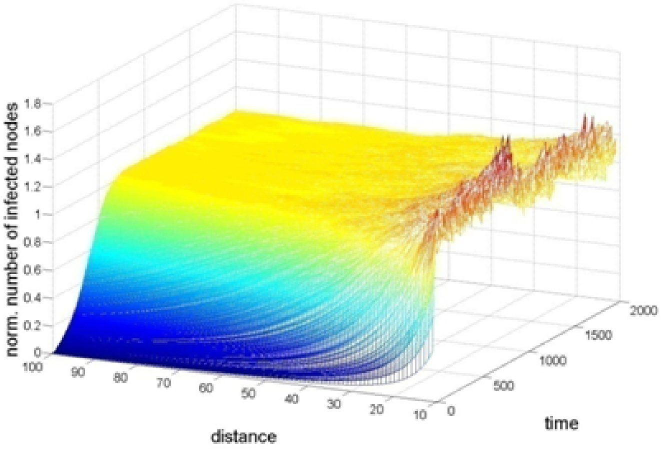


Figure 4-7: Normalized simulation results of the Validation model accounting for distance and time. Number of simulated nodes $n = 1000$ and results are averaged over 100 simulation rounds. Number of nodes at certain distance is accumulated for all distances within the distance studied. It can be seen that there is a minimum time limit needed to send information to further distances while for large time limits the dissemination realizes the full information awareness.

Moreover, the spreading is not feasible for low densities as it can be seen for low densities at Figure 4-5 which we found for low transition probability at Figure 3-8 as well.

We have calculated the normalized version of Figure 4-6, see Figure 4-7. It can be seen that after time 600, statistically all the nodes are aware of the information sent. The inaccuracy at low distances represent the higher probability of being aware of the information at cells close to the source than it would be implied by the homogeneous distribution.

We can recognize the same effect of inaccuracy at Figure 4-8 where we are showing the density-distance realization of data of the number of all nodes aware of information within a certain distance. Similar behavior has been found in Figure 3-8, namely that with lower transition probability (density) the performance of the spreading reduces significantly.

4-3 Realistic Simulations

We are calling these simulations realistic for the reason that we have used detailed street maps of Switzerland from GIS [62]. We employed three different scenarios in combination with two mobility models to conduct realistic simulations. We present and evaluate results of those realistic simulations.

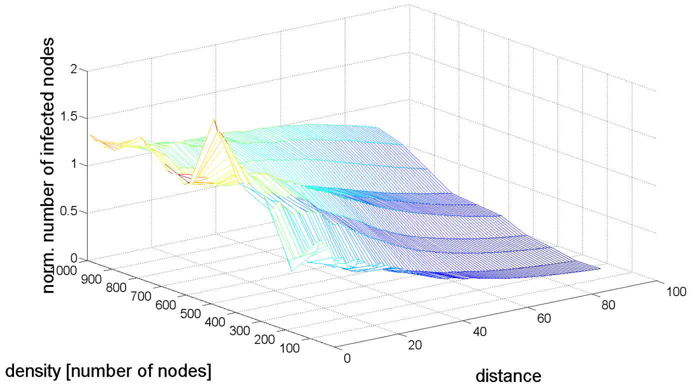


Figure 4-8: Normalized simulation results of the Validation model accounting for density and distance. Time limit $t = 1000$ and results are averaged over 100 simulation rounds. It can be seen that for smaller number of vehicles the spreading has reduced efficiency as well as for higher distances.

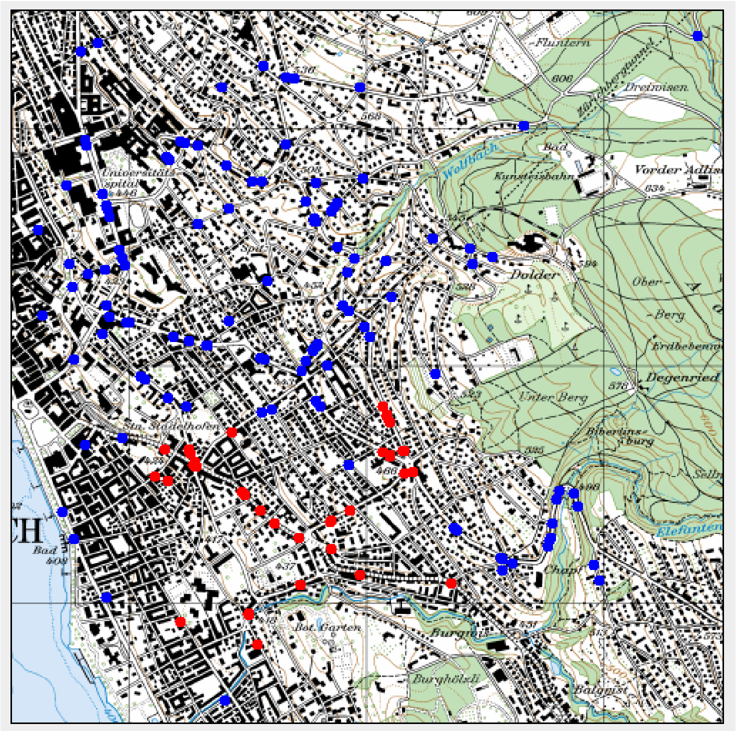


Figure 4-9: Screenshot of GMSFs simulation accounting for GIS-based road network (City scenario). Nodes colored red are aware of information while blue colored nodes are not.

Figure 4-10: City scenario for road network structure based on GIS. Downtown of city Zürich.

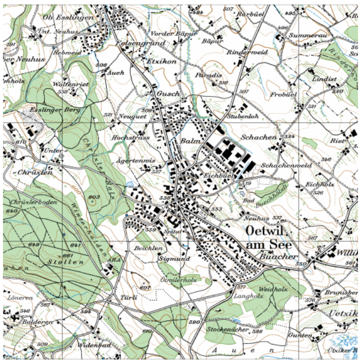


Figure 4-11: Rural scenario for road network structure based on GIS.

4-3-1 Real-World Road Structures for Realistic Simulations

In order to represent different scenarios of the road network we were using three different scenarios:

- City: Centrally located area of the city of Zürich as shown in Figure 4-10, to represent a dense road network with large number of intersections and the longest length of all roads. Therefore, it has the lowest effective density of vehicles (due to constant number of vehicles).
- Rural: Sparse road network with few intersections, see Figure 4-11, and the highest effective density of vehicles from the three scenarios studied in this work.
- Urban: This scenario is located at the board of the city of Zürich, see Figure 4-12. It is an intermediate step between the *City* and *Rural* scenarios from most point of view.

Figure 4-12: Urban scenario for road network structure based on GIS.

4-3-2 Mobility Models for Realistic Simulations

We have conducted our studies by applying two different vehicular mobility models.

Interaction-free mobility

Vehicles are traveling on the road network without interacting (e.g., avoidance of collision or adjusting speed to fore running vehicle). Their movement is only based on their constant speed and a stochastic process controlling which direction to take to exit from an intersection.

Realistic vehicular mobility

Movement model of vehicles is based on interaction among vehicles by employing realistic vehicular mobility models provided by the GMSF application (e.g., Car Following Model: deceleration to avoid collision with fortrunning vehicle; traffic lights at intersections controlled by the Intersection Management model). Vehicles are traveling at the next time moment with a speed depending on their current speed, desired speed, distance to the fore running vehicle and traffic sign at intersection ahead.

4-4 Discussion

4-4-1 Simulations in the Discussion

We have conducted our studies by applying two vehicle mobility models in combination with three scenarios and we studied the results together with the results of the Validation model (based on lattice structure and Manhattan Mobility). The combination of mobility models and road structures studied are shown in Table 4-2.

Table 4-2: List of simulations

Vehicular Mobility Model	Structure of Road Network
Manhattan Mobility	Lattice
Interaction-free	GIS: City, Urban, Rural
Realistic interactions	GIS: City, Urban, Rural

4-4-2 Expectations

Before conducting our simulations we have gathered our theoretical expectations and we are presenting them as follows:

- Vehicles will meet most frequently at the *City* scenario, than at the *Urban* scenario and with the lowest frequency at the *Rural* scenario due to the difference in the length of all roads.
- In case of Intersection Management (e.g, traffic lights) applied the high number of intersections at *City* scenario will have the highest negative effect on the dynamics of information spreading compared to the *Rural* and *Urban* scenario.
- Car Following Model will increase the level of clustering and therefore decrease the efficiency of information spreading.
- *Urban* model seems to be an intermediate step between *City* and *Rural* scenario for all the points introduced above, because its intermediate length of all roads and number of intersections.

4-4-3 Comparison of the Three Road Scenarios

We found that the spreading is the fastest at the *City* scenario, because vehicles are meeting the most frequently although the total length of the road network is higher than at other scenarios which results sparser traffic condition. Therefore, dissemination process is the fastest in case of the *City* scenario and slowest for the *Rural* scenario. Urban scenario shows an intermediate spreading compared to the *City* and *Rural* scenario.

4-4-4 Comparison of Mobility Models of Realistic Simulations

We have recognized higher level of clustering in case of applying Car Following Model and Intersection Management (e.g., traffic lights at intersections) compared to the interaction-free, stochastic-based mobility model. Clustering occurs more probably for the scenario of *City* than at *Urban* or *Rural* due to the high number

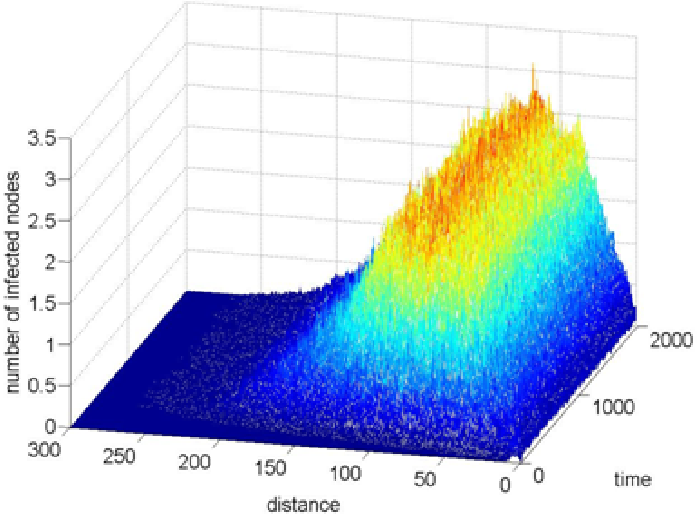


Figure 4-13: Results of GIS-based simulation accounting for interaction-free mobility model: shows number of nodes aware of information for a certain time and distance variable. Results are averaged over 100 simulation rounds.

of intersections. Figure 4-13 shows results of *City* scenario and interaction-free mobility model employed. In contrast, Figure 4-14 shows for the same *City* scenario a reduced efficiency of dissemination in case of realistic mobility models were applied.

For instance, in case of realistic vehicular mobility as shown in Figure 4-14 for distance $d = 100$ and $t = 100$, there were no nodes aware of the information while Figure 4-13 shows that nodes are more probable to be aware of the information if interaction-free mobility was applied.

4-4-5 Comparison of Realistic and Manhattan Mobility Model-Based Simulations

Comparison shows similar behavior of the simulation result for Manhattan Mobility model as shown in Figure 4-5 and simulation of GIS-based road structure in combination with interaction-free mobility, see Figure 4-13, but we found significant reduction in efficiency of information dissemination for the most realistic model shown in Figure 4-14.

We would like to note that the difference in the notation of distance does not allow us to show numerical comparison between the simulations of *validation* model and the realistic simulations. In case of realistic models the deployment of the information source is random compared to the *validation* model (center position) which leads to an extensive variation of the maximum of distance in case of realistic models

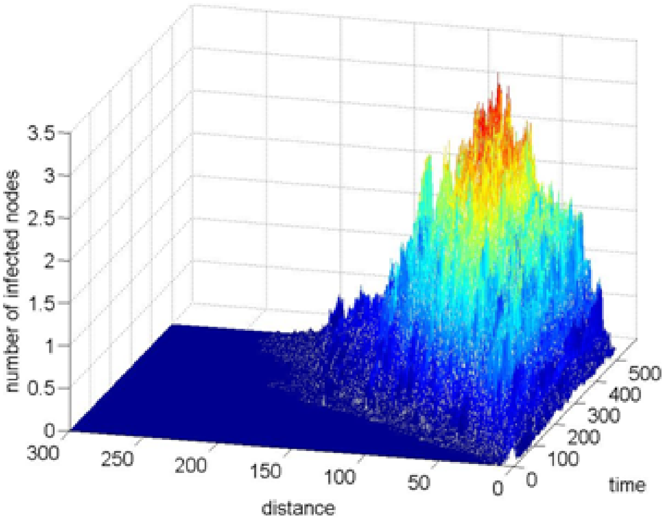


Figure 4-14: Results of GIS-based simulation accounting realistic vehicular mobility models: shows number of nodes aware of information for a certain time and distance variable. Results are averaged over 100 simulation rounds and are directly comparable with results shown in Figure 4-13 due to notation of variables is equivalent.

employed. This can be seen on Figure 4-14 and Figure 4-13 for distances $d > 100$ while the maximum for the Validation model is $d = 100$.

More significant similarity can be seen if the data is being represented at the way of taking the number of nodes aware of information within a certain distance, see Figure 4-6 and Chapter B.

4-5 Conclusion

GMSF simulator application was introduced as well as extensions of it which we implemented in order to support single source epidemic information dissemination. We have shown the connection between the *validation* model employing Manhattan Mobility and the lattice-based analytical model we introduced, studied and approximated with a mathematical model in Chapter 3.

We presented the difference of information spreading accounting for different real-world road structures and also discussed the effect of employing various mobility models (e.g., interaction-free or realistic).

Moreover, we found that spreading efficiency of realistic models is lower due to the clustering effect found, but fundamentally similar. Therefore, we believe that our analytical model with a factor accounting for the reduced efficiency can be applied to study of real-word information dissemination.

Chapter 5

Dissemination Strategy

We are presenting in this chapter the requirements of *vehicular applications*. We are describing two case studies: first to show how results of our simulations can optimize information dissemination by changing the dissemination strategy. The second case study extends the tradeoff and introduces a technique to study if resources are being wasted and propose a technique to mitigate it. We conclude our study by presenting the general dissemination strategy.

5-1 Requirements of Vehicular Applications

Different type of *vehicular applications* are having different requirements. Some of them are aiming to send messages to short distances as fast as possible while another services can provide their service at a satisfactory level by using a significantly slower information spreading in time due to the fact that their service is less time-critical. Requirement on the size of service area of applications is depending on how relevant is the information at a certain distances. Therefore, different services might have different size of area of interest and therefore they have different requirement on the distance the message have to travel.

Although *vehicular applications*'s requirements vary, we find similarities such as applications are time dependent and their services are distance dependent. Moreover, vehicles are moving on top of a well known road structure and following movement rules which are interpreted as patterns of movement.

Based on the findings above we are introducing the three requirements of *vehicular applications*:

- time (t): time needed to send an information
- distance (d): distance from the source of information
- network condition (*nc*): traffic conditions and dissemination strategy based requirements

5-1-1 Limits of Requirements

Further, we study the dependencies of the three *vehicular application* requirements as follows:

- time (t): application and its service dependent. For example, an early time accident warning system would require the fastest message transferring possible while an application for road traffic information spreading would not require that low delays to perform satisfactory.
- distance (d): like time, distance depends on the application. For instance information of an early warning system about accidents would be interesting for vehicles at close distance while information of traffic condition ahead is interesting at greater distances as well (better route planning to the destination is possible).
- network condition (nc): depends on environment dependent factors like intensity of road traffic as well as the type of road (highway, city) and direction of vehicle movements (one- bi-directional). Another factor is the technology and dissemination strategy used. There can be a great difference found for example in the transmission range depending on the communication technology used.

Further studying *vehicular applications's* requirements revealed that the applications have limits on these requirements. These can be both a maximum and minimum or just one, but at least one limit exists. An early warning system of accidents would have maximum limits on the distance (limit of interest) and a maximum allowed delay in time (in order to be able to react to the situation). The analysis, see Chapter 3, suggests that these factors are not independent from each other: by restricting the maximum distance we would free resources which would effect the choice of dissemination strategy used. In case of traffic information dissemination the distance would have been greater than previously and the maximum of distance and delay as well. However, dissemination has to be scalable with the increase of distance to keep capacity used the lowest possible (shared medium). Therefore, the frequency of relaying a message would need to be reduced (and/or aggregated) in order to keep the application scalable which is feasible due to the fact that the change (variation) of road traffic condition is low in time (readers from Holland would certainly argue this). Therefore, we investigate the limit of network condition based on distance, time limit and maximum frequency of message transmission in order to keep capacity used within the level of resources available.

5-2 Tradeoff

We present the information dissemination tradeoff with the help of two case studies where we show the direct connection between the *vehicular applications's* requirements. The analysis is, for both cases, based on results of simulations presented in Chapter 4. At first step, we show possible choices to optimize dissemination strategy to meet requirements (Case Study 1). Further, we use another representation of the same data to present the second case where we show that resources are being wasted (Case Study 2) and we propose a technique to mitigate the effect.

5-2-1 Case Study 1: Tradeoff of Time, Distance and Network Condition for a Road Traffic Information Application

We introduce the first case as an application for road traffic information dissemination which has the following requirement:

- t=600, d=80 and all the vehicles in the service area has to be aware of the information (full information awareness)

and we are aware of the fact that the number of vehicles is n= 500.

Based on simulation results introduces in Chapter 4 we found that less than 50% of all vehicles would get the information ($t = 600, d = 80$) as shown in Figure 5-1. Therefore, one of the requirement (full information awareness) is not met. Therefore, we introduce three simple ways of optimization:

- Extending the time limit (proofing that the information delay can be allowed to take longer in time): by doubling the time limit and keeping the same service area, the system is fulfilling the full information awareness criteria ($t = 1200, d = 80$).
- Reducing the size of service area by limiting distance: full awareness criteria could be fulfilled by reducing the maximum of distance to the half of the original one ($t = 600, d = 40$).
- Employing a different wireless communication technology with a significantly higher transmission range would result a faster message spreading and therefore it could fulfill all requirements stated. However, we would like to note that we show a tradeoff again, the tradeoff of the technology used. This option belongs to the tradeoff category of changing network condition.

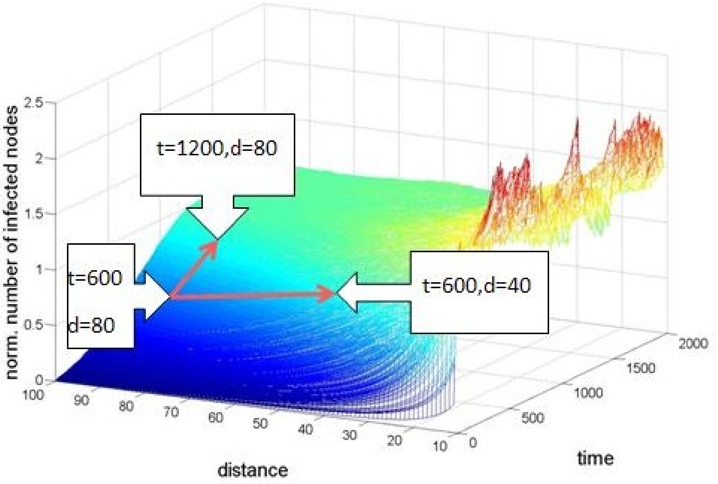


Figure 5-1: Case study of a Road Traffic Information application shows the tradeoff between Vehicular Application's requirements. The application intends to have full information awareness (of all participants) for variables ($t = 600, d = 40$). Simulation results shows the full awareness criteria not met, optimization needed: increasing time limit to $t = 1200$ or decrease distance $d = 40$.

The choices presented can be combined. Therefore we have a high level of freedom of finding the optimal solution based on the *vehicular applications's* requirements.

5-2-2 Case Study 2: Presenting Case of Wasting Radio Resource with a Road Condition Information Application

We introduce the second case of an application for disseminating road condition information. It has the following requirements:

- $t = 2000, d = 80$ and all the vehicles in the service area has to be aware of the information (full information awareness)

and we are aware of the fact that at this particular case the number of vehicles are $n = 225$.

We can see from Figure 5-2 that in case we have $n = 225$ vehicles participating in the dissemination process all the criteria are fulfilled since at $t = 2000, d = 80$ the normalized number of infected nodes is one (full information awareness). Moreover, we can see that with a third of the number of vehicles ($n = 75$) the dissemination would have the same efficiency. We would like to note that we use shared medium where we always intend to use the least amount of radio capacity. In this case we find indirect that we waste capacity by sending messages with a higher number of vehicles than needed. We propose to reduce the frequency of message sending to

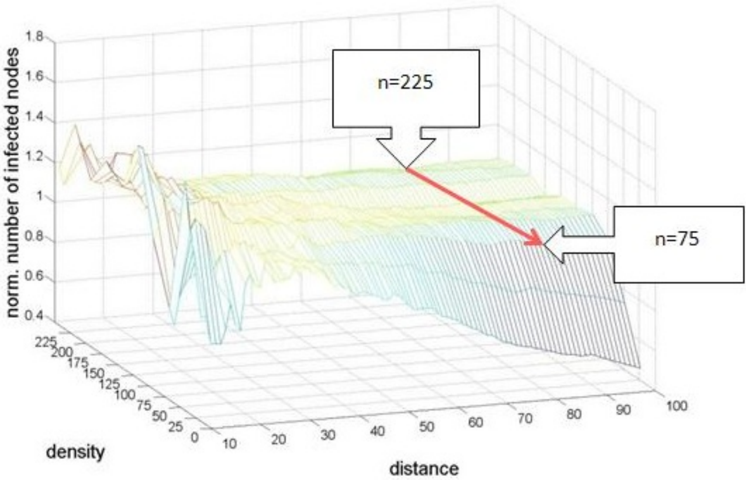


Figure 5-2: Case study of a Road Condition Information dissemination: simulation results shows for $n = 225$ and $n = 75$ number of vehicles transmitting information the full information criteria met, but radio resource is being wasted for $n = 225$. Thus, probabilistic epidemic dissemination proposed to reduce number of transmission and save resources.

match to the level of $n = 75$ which is the lowest value of n where the full information criteria stand still. Therefore, we introduce a probabilistic dissemination strategy where vehicles would relay a message according to a given probability. At this study case this probability need to be $p = 1/3$ since $1/3$ of the number of message transmission would result the same efficiency of $t = 2000$ and $d = 80$. We call a system implementing the probabilistic process introduced above a system employing probabilistic epidemic dissemination. In the second use case presented, with using probabilistic epidemic dissemination, 66% of the radio capacity would be saved in contrast to the system without this function (simple epidemic dissemination).

We present below the dissemination tradeoff for case number two as follows:

- employing probabilistic dissemination and minimizing capacity wasted
- extending maximum of distance until criteria of full awareness still stands
- reducing the time limit
- employing a different wireless communication technology with a lower level of resources needed (like energy), smaller transmission range and lower level of signal interference.

5-2-3 Impact of Traffic Conditions Variation

We introduce the daily variation of road traffic. Studies are showing that the number of vehicles vary by the hour of day studied, see [54] [65] [66]. This fact has a

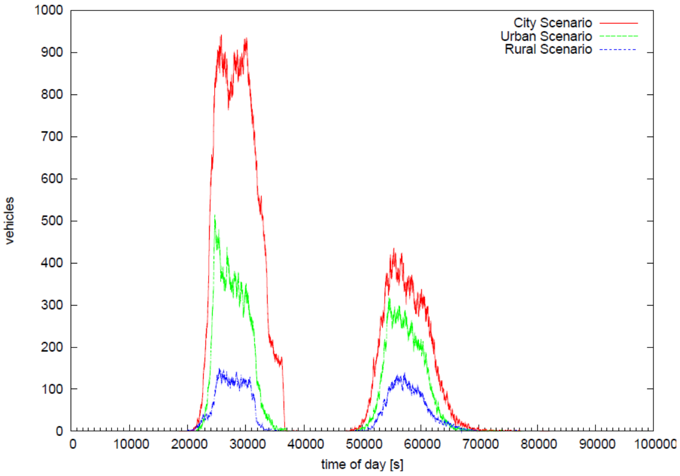


Figure 5-3: Measurement results of variation of daily traffic

remarkable influence on our studies though we presented that network condition is depending on the road traffic, number of vehicles participating in the information dissemination. Therefore, we propose to extend the dissemination tradeoff accounting for traffic variation. Figure 5-3 from [54] shows the variation of number of vehicles for the three real-world road scenarios we have studied at Chapter 4. Variation of traffic results dramatic performance variation as shown in [66], which implies that application parameters are influenced by the hour of day they are studied and intended to provide its service.

Therefore, we propose to study the usage of applications with not continuously available services. An example use case is a *vehicular application* for road traffic information dissemination. We experience traffic jams at the time of day when the density of vehicles is relatively high (leads to traffic jams). Therefore, we propose an application providing its service if the number of vehicles are high enough (meeting requirements of satisfactory level of service), in which case the existence of traffic jams are highly probable and information is valuable.

5-2-4 Conclusion

We have introduced the three requirements of *vehicular applications* (time, distance and network condition) which are depending on application limits, traffic conditions and the dissemination strategy considered. These variables are depending on each other (simplified representation can be seen at Figure 5-4) We can find the optimum by changing them without exceeding limits which make applications unfeasible to use. We call these additional limits threshold levels. For example, an application's maximum time limit for disseminating road condition information can vary between

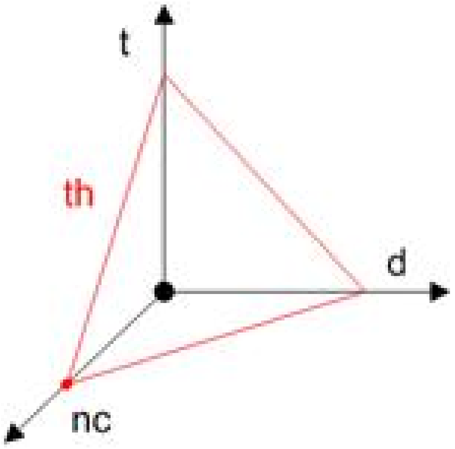


Figure 5-4: Tradeoff of the three vehicular application's requirements and their threshold values represented as a surface (example case)

couple of kilometers up to a distance where the road conditions are changing significantly, which we consider the threshold level of such an application.

5-3 Summary of Dissemination Strategy

We have introduced the three requirements of *vehicular applications* (time, distance and network condition) and we presented the connection among them by showing the tradeoff of information dissemination.

We showed in *Case Study 1* the possible optimization of dissemination and its rewards. We proposed in *Case Study 2* a probabilistic epidemic dissemination to reduce wasted capacity, increase efficiency and we proposed applications which are providing their services at limited, but most important, time of day by taking advantage of studies showing variance of road traffic by the hour of day.

In case the dissemination optimization leads to a strategy which does not exceed any threshold levels of the *vehicular application's* requirements the epidemic information dissemination are recommended to be used.

We further study the case if the dissemination optimization cannot lead to a result which is not exceeding the previously introduced threshold level. Surpassing threshold levels are not acceptable, because these are based on fundamental or elementary limitations. In order to be able to provide sufficient level of service we propose to use additional infrastructure-based wireless communication like the cellular-based 3G system which has an extensive availability area and does not depend on the vehicles itself.

We summarize our findings as follows: Epidemic dissemination is recommended with optimization of dissemination strategy. In case of *vehicular application's*s requirements are surpassing thresholds, we propose to use infrastructure based communication technology (e.g., 3G). Contrarily, if we find resources being wasted, we suggest to employ the probabilistic dissemination introduced in Section 5-2-2.

Chapter 6

Conclusion

6-1 Conclusions

In this work we introduced an analytical model based on homogeneous transition probability and lattice-based structure in order to study information spreading in *vehicular networks*. We introduced a General Recursive model, its simplified version, the Recursive First-wave model, to analyze information spreading extensively. We have shown theoretically that these models overestimate the reality. Therefore, we have introduced the Compensated model to tackle the over estimation based on the wrong estimation of maximum number of vehicles at different distances. We presented upper and lower bound approximation of the *first-wave* case and showed their relation to the models introduced before.

We have evaluated the analytical models and approximations with results of numerical simulations. We studied the behavior of information dissemination based on these numerical simulation's results and compared the *general* and *first-wave* cases. We found similarities and differences of the two case and defined a *phase transition* for the *first-wave* case. We introduced a mathematical model to approximate results of numerical evaluations and found that this model fits only for transition probabilities lower than *phase transition*. We extended successfully our model to approximate well for transition probabilities higher than *phase transition*. We presented and evaluated the results based on the results of the numerical simulations of the *first-wave* case.

Simulations in our work were aiming to find the connection between numerical results and realistic simulations with various levels of reality. Simulations were performed with a realistic simulator application providing microscopic simulation of different vehicular mobility models of vehicles. In order to support simulation of epidemic type of information dissemination, we have extended capability of the simulator application with the necessary functions. We have shown the similar behavior of dissemination for the numerical simulation results of the lattice-based, analytical model we defined and Manhattan Mobility model, simulated by the simulator application. Simulation of entity-based mobility with real-world based road network structures showed small difference from the lattice based mobility model while we found high level of clustering for realistic vehicular mobility models. Clustering of vehicles were more significant for the *City* scenario than compared to the *Rural* and *Urban* scenario. This is a negative effect which has reduced efficiency of information dissemination.

We introduced the dissemination tradeoff and showed the optimization of the dissemination strategy based on the simulation results with the help of a use case. With the help of a second use case, we introduced a technique to find resources wasted by a *vehicular application* and for that case, we proposed to implement the probabilistic epidemic dissemination.

Finally, we showed the dissemination strategy tradeoff. We presented the requirements of using Vehicle-to-Vehicle communication as employing optimized, *epidemic* type of dissemination. In addition, we proposed to use *probabilistic epidemic* dissemination if resources being wasted. Eventually, we presented the need for *hybrid technology*, employing infrastructure-based technology (like 3G) as well, for cases the dissemination optimization exceeds fundamental and elemental limits of *vehicular application's* requirements.

Our study with the help of the simulator application can be used to prevent early stage, on-site experiments aiming to test if a *vehicular application* could provide satisfactory level of service for a certain application scenario. With the help of our work, simulations and analysis can show if requirements of a *vehicular application* can be met or has to be optimized and how the dissemination strategy tradeoff need to be applied.

6-2 Future Work

We aim to extend the analytical model introduced in Chapter 3 to a general closed-form model and find the physical meaning of it. Further improvement of mapping reality by extending the vehicular mobility model is needed as well as more detailed simulations about the variation of traffic by time of day.

Appendix A

Most Important Source Codes

A-1 Numerical Simulation of Lattice Model

```
function lattice = makeLatticeProb(edge,maxdistance,lattice,p)
    %calculating the center and edge distance
    center = maxdistance +1;
    %edge = 2* maxdistance + 1;

    %initialization of the temporary lattice and center node
    lattice(center,center)=1;
    latticetemp=lattice;

    for i=1:maxdistance

        for z=0:1:(i-1)
            if (center+z+1 <= edge && center-i >= 1)

                if lattice(center-i+z,center+z) == 1
                    if rand(1) >= 1-p
                        latticetemp(center-i+z+1,center+z) = 1;
                    end

                    if (center-i+z-1 >= 1 && rand(1) >= 1-p)
                       latticetemp(center-i+z-1,center+z)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center-i+z,center+z+1)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center-i+z,center+z-1)= 1;
                    end
```

```
                end
            end
        end

        for z=0:1:(i-1)
            if (center+i <= edge && center+1-z >= 1 && center+z <= edge)

                if lattice(center+z,center+i-z) == 1
                    if rand(1) >= 1-p
                        latticetemp(center+z+1,center+i-z)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center+z-1,center+i-z)= 1;
                    end

                    if (center+i-z+1 <= edge && rand(1) >= 1-p)
                        latticetemp(center+z,center+i-z+1)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center+z,center+i-z-1)= 1;
                    end
                end
            end
        end

        for z=0:1:(i-1)
            if (center+i <= edge && center-z >= 1)

                if lattice(center+i-z,center-z) == 1
                    if (center+i-z+1 <= edge && rand(1) >= 1-p)
                        latticetemp(center+i-z+1,center-z)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center+i-z-1,center-z)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center+i-z,center-z+1)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center+i-z,center-z-1)= 1;
                    end
                end
            end
        end

        for z=0:1:(i-1)
            if (center-i+z <= edge && center-z >= 1 && center-i >= 1)

                if lattice(center-z,center-i+z) == 1
```

```
                    if rand(1) >= 1-p
                        latticetemp(center-z+1,center-i+z)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center-z-1,center-i+z)= 1;
                    end

                    if rand(1) >= 1-p
                        latticetemp(center-z,center-i+z+1)= 1;
                    end

                    if (center-i+z-1 >= 1 && rand(1) >= 1-p)
                        latticetemp(center-z,center-i+z-1)= 1;
                    end
                end
            end
        end

    end

    %Nodes at distance 1
    if rand(1) >= 1-p
        latticetemp(center+1,center)= 1;
    end

    if rand(1) >= 1-p
        latticetemp(center-1,center)= 1;
    end

    if rand(1) >= 1-p
        latticetemp(center,center-1)= 1;
    end

    if rand(1) >= 1-p
        latticetemp(center,center+1)= 1;
    end

    %lattice gets updated with all nodes are being infected at the current
    %time moment
    lattice = lattice | latticetemp;

end
```

A-2 Numerical Simulation of the First-wave Case for Lattice Model

```
function [NFristWave,variance] = firstWaveSimulation( probresolution, timedistancemaximum,numberofsimulationrounds,v )
    NFristWave=zeros(probresolution,timedistancemaximum);
    variance=zeros(probresolution,timedistancemaximum);

```

```
    for e=1:timedistancemaximum
        [NFristWaveTemp,X]=simulation(probresolution,e,e,numberofsimulationrounds,v);

        for c=1:size(NFristWaveTemp,1)
           NFristWave(c,e)= NFristWaveTemp(c,e);
        end

        if v==1
            for c=1:size(NFristWaveTemp,1)
                variance(c,e)= X(c,e);
            end
        end

    end
end
```

A-3 Calculation of the Number of Nodes Aware of Information for the Lattice Model

```
function [N,variancePerProb] = simulation(probresolution,maxdistance,maxtime,numberofsimulationrounds,v)
    probstep = 1 / probresolution;
    N = zeros(probresolution,maxdistance);
    variancePerProb=zeros(1,maxdistance);

    %initialization:making a matrix as a reference for distances
    if maxdistance ==1
        edge=3;
    else
    edge = 2 * maxdistance + 1;
    end

    latticeDistance=zeros(edge,edge);
    latticeDistance = makeLatticeDistance(maxdistance);

    for q=probstep:probstep:1
        %initialization
        numberofnodesatdistancesSUM=zeros(1,maxdistance);

        varianceperprob=zeros(numberofsimulationrounds,maxdistance);

        %rounds of simulations
        for l = 1:numberofsimulationrounds
            %initialization of lattice grid
            lattice = zeros(edge,edge);

            %infected lattices after maxtime
            for i=1:maxtime
                lattice = makeLatticeProb(edge,maxdistance,lattice,q);
            end

```

```
            %initialization:matrix to store the number of infected nodes at a
            %certain distance from center node
            numberofnodesatdistances=zeros(1,maxdistance);

            %checking all lattice point and updating the matrix which stores
            %the number of infected nodes at a certain distance
            for j=1:edge
               for k=1:edge
                   if  (lattice(j,k) == 1 && latticeDistance(j,k) ~= 0)
                    numberofnodesatdistances(1,latticeDistance(j,k)) = numberofnodesatdistances(1,latticeDistance(j,k)) + 1;
                   end
               end
            end

            %adding numberofnodesatdistances to a varianceperpro
            if v==1
                varianceperprob(1,:) = numberofnodesatdistances;
            end


            %summs up all the number of infected nodes at a certain distance
            %after several simulation round
            numberofnodesatdistancesSUM= numberofnodesatdistancesSUM + numberofnodesatdistances;
        end

        %normalizing with the number of rounds
        b=round(q*probresolution);
        for n=1:maxdistance
            N(b,n)= numberofnodesatdistancesSUM(1,n)/numberofsimulationrounds;
        end

        %put the variance data from different probabilities together
        if v==1
            variancePerProb(b,:)=var(varianceperprob,0,1);
        end
    end
end
```

Appendix B

Additional Results of Simulations

B-1 Simulation Results of the Validation Model in GMSF

B-2 GIS-based Simulations with GMSF

B-3 Numerical simulations of the General model

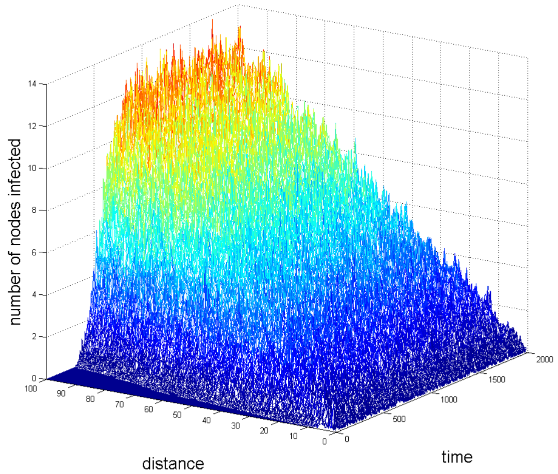


Figure B-1: Simulation result of the Validation model based on lattice road structure and Manhattan Mobility shows the number of nodes aware of information for certain distance and time variables (not accumulated for distance). Number of simulated nodes is $n = 1000$ and results are averaged over 100 simulation rounds. Spreading for high distances have a minimum requirement on time limit.

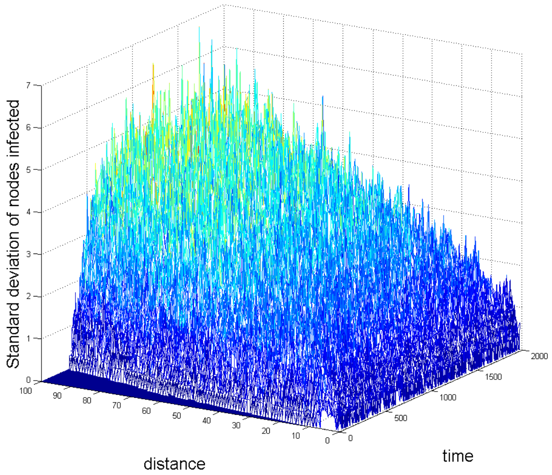


Figure B-2: Standard deviation of simulation result of the Validation model based on lattice road structure and Manhattan Mobility shows the number of nodes aware of information for certain distance and time variables (not accumulated for distance). Number of simulated nodes is $n = 1000$ and results are averaged over 100 simulation rounds. Spreading for high distances have a minimum requirement on time limit.

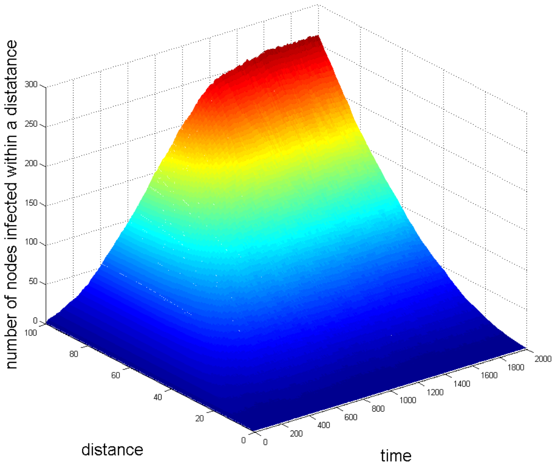


Figure B-3: Simulation result of the Validation model based on lattice road structure and Manhattan Mobility shows the fraction of nodes aware of information for certain distance and time variables (accumulated for distance). Number of simulated nodes is $n = 500$ and results are averaged over 100 simulation rounds. Spreading for high distances have a minimum requirement on time limit.

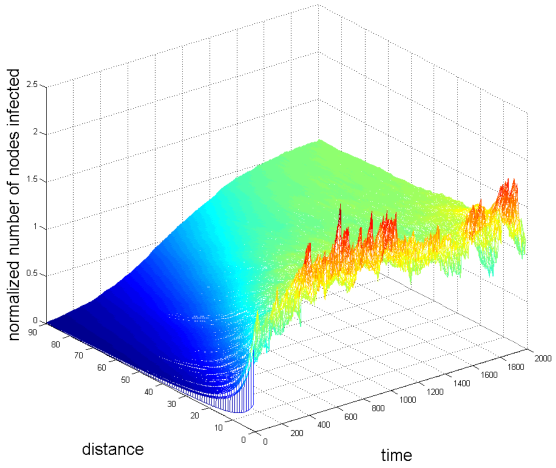


Figure B-4: Normalized simulation result of the Validation model based on lattice road structure and Manhattan Mobility shows the fraction of nodes aware of information for certain distance and time variables (accumulated for distance). Number of simulated nodes is $n = 400$ and results are averaged over 100 simulation rounds.

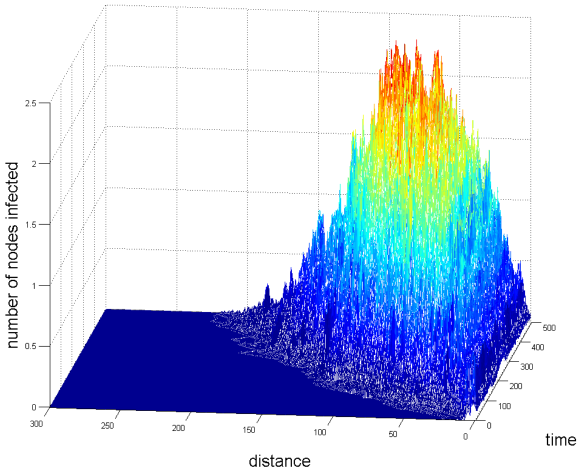


Figure B-5: Results of GIS-based, City scenario, simulation accounting realistic vehicular mobility models: shows number of nodes aware of information for a certain time and distance variable. Number of vehicles $n = 200$ and the results are averaged over 100 simulation rounds.

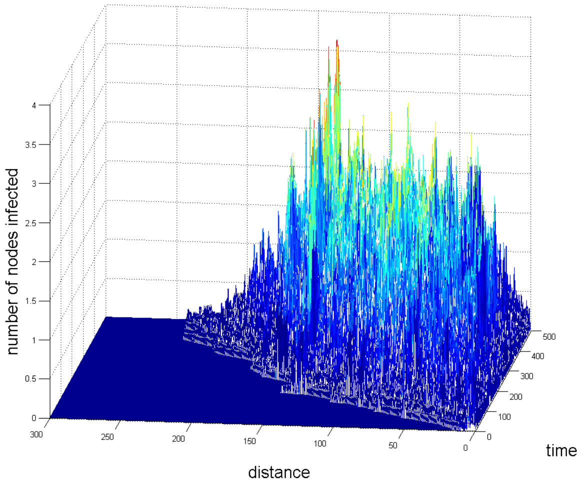


Figure B-6: Results of GIS-based, Urban scenario, simulation accounting realistic vehicular mobility models: shows number of nodes aware of information for a certain time and distance variable. Number of vehicles $n = 200$ and the results are averaged over 100 simulation rounds.

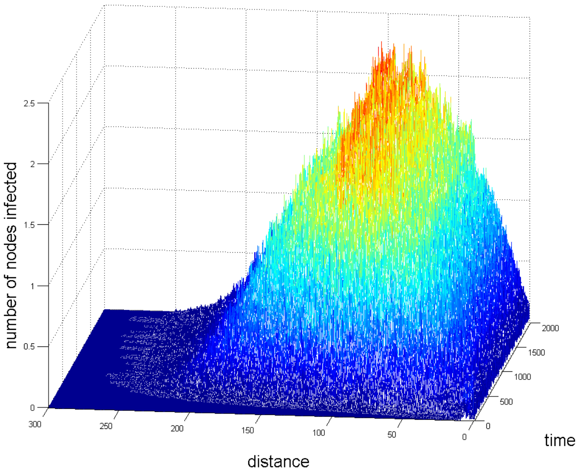


Figure B-7: Results of GIS-based, City scenario, simulation accounting for interaction-free vehicular mobility: shows number of nodes aware of information for a certain time and distance variable. Number of vehicles $n = 200$ and the results are averaged over 100 simulation rounds.

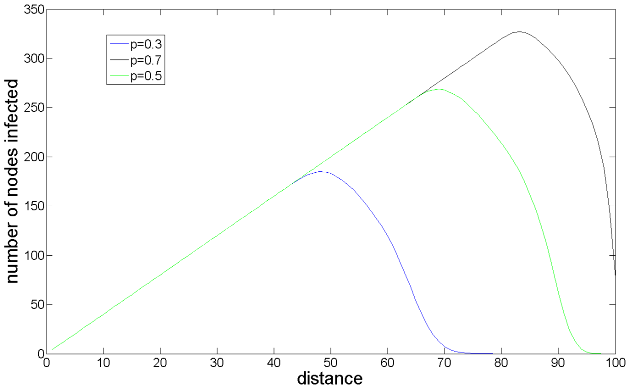


Figure B-8: Numerical simulation result of the General model accounting for distance and various transition probabilities shows the number of nodes aware of information for certain distance (not accumulated for distance). Number of simulated nodes is $n = 1000$ and results are averaged over 100 simulation rounds. Spreading for higher distances requires a higher transition probability.

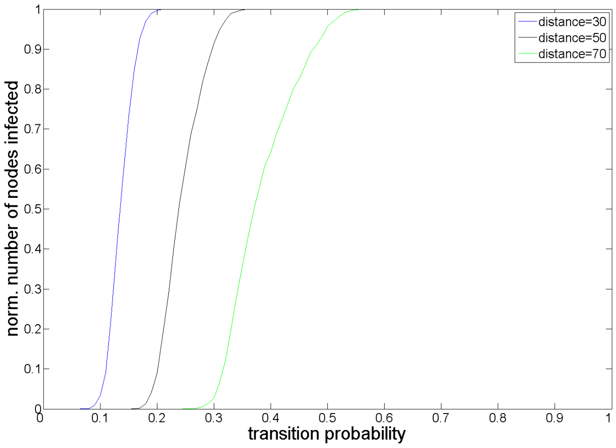


Figure B-9: Numerical simulation result of the General model accounting for transition probability and various distances shows the normalized number of nodes aware of information for certain distance (not accumulated for distance). Number of simulated nodes is $n = 1000$ and results are averaged over 100 simulation rounds. Spreading for higher distances requires a higher transition probability.

Bibliography

[1] L. Wischoff, A. Ebner, H. Rohling, M. Lott, and R. Halfmann, "Sotis, a self-organizing traffic information system," in *IEEE Vehicular Technology Conference (VTC)*, 2003, pp. 2442 – 2446.

[2] D. Valerio, A. D'Alconzo, F. Ricciato, and W. Wiedermann, "Exploiting cellular networks for road traffic estimation: a survey and a research roadmap," in *IEEE Vehicular Technology Conference (VTC)*, 2009, pp. 1–5.

[3] J. Yoon, B. Noble, and M. Liu, "Surface street traffic estimation," in *ACM international conference on Mobile systems, applications and services*, 2007, pp. 220–232.

[4] B. Yu, J. Gong, and C. Xu, "Catch-up: a data aggregation scheme for vanets," in *ACM international workshop on Vehicular Internetworking*, 2008, pp. 49–57.

[5] R. Lu, X. Lin, H. Zhu, and X. Shen, "Spark: a new vanet-based smart parking scheme for large parking lots," in *IEEE INFOCOM*, 2009, pp. 1413–1421.

[6] M. Maile, "Vehicle safety communications-applications (vsc-a) project: Crash scenarios and safety applications," in *SA Goverment/Industry Meeting*, 2009.

[7] W. Kermack, "A contribution to the mathematical theory of epidemics," in *Proc. R. Soc. Lond. A*, 1927, pp. 700–721.

[8] F. Brauer and C. Castillo-Chávez, *Mathematical models in population biology and epidemiology.* Springer Verlag, 2001.

[9] D. Daley, J. Gani, and J. Gani, *Epidemic modelling: an introduction.* Cambridge Univ Press, 2001.

[10] P. Grassberger, “On the critical behavior of the general epidemic process and dynamical percolation,” *Mathematical Biosciences*, pp. 157–172, 1983.

[11] J. Cardy and P. Grassberger, “Epidemic models and percolation,” *Journal of Physics A: Mathematical and General*, p. L267, 1985.

[12] H. Kesten, *Percolation theory for mathematicians.* Birkhäuser, 1982.

[13] I. Jensen, “Low-density series expansions for directed percolation on square and triangular lattices,” *Journal of Physics A: Mathematical and General*, p. 7013, 1996.

[14] R. Albert and A. Barabási, “Statistical mechanics of complex networks,” *Reviews of modern physics*, 2002.

[15] X. Xu, Z. Wu, and G. Chen, “Epidemic spreading in lattice-embedded scale-free networks,” *Physica A: Statistical Mechanics and its Applications*, pp. 125–130, 2007.

[16] P. Wang and M. González, “Understanding spatial connectivity of individuals with non-uniform population density,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, p. 3321, 2009.

[17] S. Dashtinezhad, T. Nadeem, B. Dorohonceanu, C. Borcea, P. Kang, and L. Iftode, “TrafficView: a driver assistant device for traffic monitoring based on car-to-car communication,” in *IEEE, Vehicular Technology Conference(VTC)*, 2004.

[18] Y. Zhuang, J. Pan, and L. Cai, “A probabilistic model for message propagation in two-dimensional vehicular ad-hoc networks,” in *ACM international workshop on Vehicular Internetworking*, 2010, pp. 31–40.

[19] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, “VTrack: accurate, energy-aware road traffic delay estimation using mobile phones,” in *ACM Conference on Embedded Networked Sensor Systems*, 2009, pp. 85–98.

[20] P. Mohan, V. Padmanabhan, and R. Ramjee, “Nericell: rich monitoring of road and traffic conditions using mobile smartphones,” in *ACM conference on Embedded network sensor systems*, 2008, pp. 323–336.

[21] R. Sen, B. Raman, and P. Sharma, “Horn-ok-please,” in *ACM international conference on Mobile systems, applications, and services*, 2010, pp. 137–150.

[22] C. Lochert, B. Scheuermann, and M. Mauve, “Probabilistic aggregation for data dissemination in VANETs,” in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2007, pp. 1–8.

[23] C. Lochert, B. Scheuermann, C. Wewetzer, A. Luebke, and M. Mauve, "Data aggregation and roadside unit placement for a vanet traffic information system," in *ACM international workshop on Vehicular Internetworking*, 2008, pp. 58–65.

[24] C. Maihöfer, T. Leinmüller, and E. Schoch, "Abiding geocast: time–stable geocast for ad hoc networks," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2005, pp. 20–29.

[25] J. Rybicki, B. Scheuermann, M. Koegel, and M. Mauve, "PeerTIS: a peer-to-peer traffic information system," in *ACM international workshop on Vehicular Internetworking*, 2009, pp. 23–32.

[26] G. Korkmaz, E. Ekici, F. Özgüner, and U. Özgüner, "Urban multi-hop broadcast protocol for inter-vehicle communication systems," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2004, pp. 76–85.

[27] Q. Xu, T. Mak, J. Ko, and R. Sengupta, "Vehicle-to-vehicle safety messaging in DSRC," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2004, pp. 19–28.

[28] T. ElBatt, S. Goel, G. Holland, H. Krishnan, and J. Parikh, "Cooperative collision warning using dedicated short range wireless communications," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2006, pp. 1–9.

[29] J. Rybicki, B. Scheuermann, W. Kiess, C. Lochert, P. Fallahi, and M. Mauve, "Challenge: peers on wheels-a road to new traffic information systems," in *ACM international conference on Mobile computing and networking*, 2007, pp. 215–221.

[30] E. Giordano, R. Frank, G. Pau, and M. Gerla, "CORNER: a step towards realistic simulations for VANET," in *ACM international workshop on Vehicular Internetworking*, 2010, pp. 41–50.

[31] J. Zhao, T. Arnold, Y. Zhang, and G. Cao, "Extending drive-thru data access by vehicle-to-vehicle relay," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2008, pp. 30–39.

[32] Y. Ding, C. Wang, and L. Xiao, "A static-node assisted adaptive routing protocol in vehicular networks," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2007, pp. 59–68.

[33] T. Spyropoulos, K. Psounis, and C. Raghavendra, "Efficient routing in intermittently connected mobile networks: the single-copy case," *IEEE/ACM Transactions on Networking (TON)*, pp. 63–76, 2008.

[34] B. Scheuermann, C. Lochert, J. Rybicki, and M. Mauve, "A fundamental scalability criterion for data aggregation in VANETs," in *ACM International conference on Mobile computing and networking*, 2009, pp. 285–296.

[35] J. Haerri and C. Bonnet, "On the classification of routing protocols in mobile ad-hoc networks," Institut Eurecom, France, Tech. Rep. EURECOM+1492, 08 2004.

[36] S. Dietzel, B. Bako, E. Schoch, and F. Kargl, "A fuzzy logic based approach for structure-free aggregation in vehicular ad-hoc networks," in *ACM international workshop on Vehicular Internetworking.* ACM, 2009, pp. 79–88.

[37] M. Caliskan, D. Graupner, and M. Mauve, "Decentralized discovery of free parking places," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2006, pp. 30–39.

[38] P. Costa, D. Frey, M. Migliavacca, and L. Mottola, "Towards lightweight information dissemination in inter-vehicular networks," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2006.

[39] H. Wu, R. Fujimoto, R. Guensler, and M. Hunter, "MDDV: a mobility-centric data dissemination algorithm for vehicular networks," in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2004, pp. 47–56.

[40] A. Khelil, C. Becker, J. Tian, and K. Rothermel, "An epidemic model for information diffusion in MANETs," in *ACM international workshop on Modeling analysis and simulation of wireless and mobile systems*, 2002, pp. 54–60.

[41] F. Bai and A. Helmy, "A survey of mobility models in wireless adhoc networks," 2006.

[42] K. Pearson, "The problem of the random walk," *Nature*, pp. 294–294, 1905.

[43] G. Farhadi and N. Beaulieu, "On the connectivity and average delay of mobile ad hoc networks," in *IEEE International Conference on Communications(ICC)*, 2006, pp. 3868–3872.

[44] J. Harri, F. Filali, and C. Bonnet, "Mobility models for vehicular ad hoc networks: a survey and taxonomy," *IEEE Communications Surveys & Tutorials*, pp. 19–41, 2009.

[45] S. Heimlicher and K. Salamatian, "Globs in the primordial soup: the emergence of connected crowds in mobile wireless networks," in *ACM international symposium on Mobile ad hoc networking and computing*, 2010, pp. 91–100.

[46] M. Kerper, W. Kiess, and M. Mauve, "Coordinated VANET experiments: a methodology and first results," in *ACM international workshop on Vehicular Internetworking*, 2009, pp. 109–116.

[47] L. D. Hartog, T. Spyropoulos, and F. Legendre, "Using public transportation as a DTN backbone."

[48] M. Artimy, W. Robertson, and W. Phillips, “Assignment of dynamic transmission range based on estimation of vehicle density,” in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2005, pp. 40–48.

[49] H. Wedde, S. Lehnhoff, and B. van Bonn, “Highly dynamic and scalable VANET routing for avoiding traffic congestions,” in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2007, pp. 81–82.

[50] H. Schumacher, C. Priemer, and E. Slottke, “A simulation study of traffic efficiency improvement based on Car-to-X communication,” in *ACM international workshop on Vehicular Internetworking*, 2009, pp. 13–22.

[51] H. Hartenstein, K. Laberteaux, and E. Corporation, *VANET Vehicular Applications and Inter-Networking Technologies.* Wiley Online Library, 2010.

[52] M. Fiore and J. Härri, “The networking shape of vehicular mobility,” in *ACM international symposium on Mobile ad hoc networking and computing*, 2008, pp. 261–272.

[53] R. Baumann, S. Heimlicher, and M. May, “Towards realistic mobility models for vehicular ad-hoc networks,” in *Mobile Networking for Vehicular Environments*, may 2007, pp. 73 –78.

[54] R. Baumann, F. Legendre, and P. Sommer, “Generic mobility simulation framework (GMSF),” in *ACM SIGMOBILE workshop on Mobility models*, 2008, pp. 49–56.

[55] M. Killat, F. Schmidt-Eisenlohr, H. Hartenstein, C. Rössel, P. Vortisch, S. Assenmacher, and F. Busch, “Enabling efficient and accurate large-scale simulations of VANETs for vehicular traffic management,” in *ACM international workshop on Vehicular Ad hoc Networks(VANET)*, 2007, pp. 29–38.

[56] C. Gorgorin, V. Gradinescu, R. Diaconescu, V. Cristea, and L. Ifode, “An integrated vehicular and network simulator for vehicular ad-hoc networks,” in *European Simulation and Modelling Conference*, 2006.

[57] M. Artimy, W. Phillips, and W. Robertson, “Connectivity with static transmission range in vehicular ad hoc networks,” *Communication Networks and Services Research Conference (CNSR)*, pp. 237–242, 2005.

[58] B. Liu, B. Khorashadi, D. Ghosal, C. Chuah, and M. Zhang, “Assessing the VANET’s local information storage capability under different traffic mobility,” in *IEEE INFOCOM*, 2010, pp. 1–5.

[59] J. Mittag, F. Thomas, J. Härri, and H. Hartenstein, “A comparison of single- and multi-hop beaconing in VANETs,” in *ACM international workshop on Vehicular Internetworking*, 2009, pp. 69–78.

[60] M. Fiore, J. Harri, F. Filali, and C. Bonnet, “Vehicular mobility simulation for VANETs,” 2007.

[61] J. Härri, M. Fiore, F. Filali, and C. Bonnet, “Vehicular mobility simulation with vanetmobisim,” *SIMULATION*, p. 275, 2009.

[62] Vector 25 - landscape model of switzerland. The Swiss Federal Office of Topography. [Online]. Available: http://www.swisstopo.ch/en/products/digital/landscape/vec25/

[63] S. McCanne, S. Floyd, K. Fall, K. Varadhan *et al.*, “Network simulator ns-2,” 1997.

[64] Q. Simulator, “Scalable network technologies,” *Inc.[Online]. Available: www. qualnet. com.*

[65] B. Raney, A. Voellmy, N. Cetin, M. Vrtic, and K. Nagel, “Towards a microscopic traffic simulation of all of Switzerland,” *Computational Science ICCS*, pp. 371–380, 2002.

[66] F. Bai and B. Krishnamachari, “Spatio-temporal variations of vehicle traffic in VANETs: facts and implications,” in *ACM international workshop on Vehicular Internetworking*, 2009, pp. 43–52.

Glossary

List of Acronyms

CAN	Content Adressable Network
CFM	Car Following Model
DTN	Delay Tolerant Network
GIS	Geographic Information System
GMSF	Generic Mobility Simulation Framework
GUI	Graphical User Interface
IDM	Intelligent Driver Model
MANET	Mobile Ad-Hoc Network
ns	The Network Simulator
RMSE	Root-Mean-Square Error
RSU	Road-Side Unit
RWM	Random Walk Model
RWP	Random Waypoint Model
SI	Susceptible, Infectious
SIR	Susceptible, Infectious, Removed
SUMO	Simulator of Urban MObility
TCP	Transmission Control Protocol
TMC	Traffic Message Channel

TTL	Time To Live
VANET	Vehicular Ad-Hoc Network
VSC-A	Vehicular Safety Communication-Applications
V2V	Vehicle-to-Vehicle
XML	Extensible Markup Language
3G	3rd Generation mobile telecommunications