DISTRIBUTED TRAFFIC INFORMATION SYSTEM (TIS) BASED ON V2X COMMUNICATION IN LARGE-SCALE URBAN ENVIRONMENTS

NICHOLAS LOULLOUDES

A DISSERTATION
SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT THE
UNIVERSITY OF CYPRUS

RECOMMENDED FOR ACCEPTANCE
BY THE DEPARTMENT OF
COMPUTER SCIENCE

JUNE 2015

Abstract

Intelligent Transportation Systems (ITS) aim to assist commuters in taking informed decisions concerning travel safety and efficiency. Vehicular Ad Hoc Networks (VANET) are one of the key platforms for ITS, wherein vehicle-to-vehicle and vehicle-to-infrastructure communication - collectively designated as V2X - can support information dissemination. With high-fidelity sensors becoming mainstream in modern vehicles, VANETs are expected to adequately support the diffusion of disparate sensory artifacts, including the prevailing traffic conditions. In turn, V2X-based Traffic Information Systems (TIS) will provide drivers with dynamic route planning and congestion estimates. Nevertheless, the intrinsic properties of mobility, the lack of clear understanding of vehicular network connectivity, as well as the stringent time and quality constraints for traffic data diffusion, impose significant barriers in the process of designing, implementing and deploying TIS. Moreover, the research community has yet to adopt thorough and realistic performance evaluations approaches for such systems and their peripheral components thereby introducing unnecessary risks of failure, simply because they will confront a demanding environment they were not designed for.

This thesis will present the research efforts towards providing solutions to the aforementioned problems. To investigate the potentials and understand how distributed VANET-based TIS will eventually be realized, we study through extensive analytical and simulative methods the following problem domains: (i) VANET dynamics in Urban Environments, (ii) Traffic Sensing and Acquisition, and (iii) Large-Scale Traffic Information Dissemination. At first, we explore the basic premise behind V2X, specifically the constantly evolving spatio-temporal nature of urban-based VANETs through which traffic information will flow. Here, V2X connectivity is modeled as sequences of contact graphs wherein complex network analysis measures are employed to gain insights on the structural properties of VANETs from a city-wide perspective. In contrast to other low-dimensional networks, we unveil the VANET's high sensitivity to fragmentation, extremely variable connectivity and total lack of small-world and scale-free features. Additionally, we present why and where particular network phenomena occur in the underlying road network and exploit the implications in the design of effective information dissemination mechanisms. Next, we consider the constrained bandwidth of VANETs, and thus we investigate the possibility of in-vehicle data caching. Particularly, we examine its capacity to improve the efficiency of TIS by reducing the time-to-serve requests and network overhead, while abiding established time and quality constraints. Through large-scale realistic simulative evaluations, we discuss that the use of simple TTL-based replacement policies can achieve significant improvements under both normal traffic conditions and unscheduled traffic events. Finally, we present V-Radar a reactive, VANET-based, traffic information dissemination protocol for metropolitan environments. In contrast to other approaches in the literature, V-Radar enables the querying and acquisition of traffic information along a number of composite road-paths, starting from a vehicle's current position towards its final destination. Ultimately, this effective approach enables the driver to establish a more broad view of the traffic conditions that will be encountered further ahead in a timely manner. Finally, we present the analytical and simulation testbeds developed to support the above research, which are open-source released to the community for future studies.

Acknowledgements

Years ago, someone wisely told me: *"Enrolling into a PhD programme is similar to being dropped in the middle of an ocean; your job is to find the way to the closest land, and make it their alive"*. Indeed, this endeavour for me was a rather long swim home, a swim of 8 years with its share of ups and downs. However I always felt the support of several instrumental people throughout this journey that each one propelled me forward in their own way. Here I would like to acknowledge them.

First and foremost, I would like to wholeheartedly express my sincere gratitude to my academic supervisor Professor Marios D. Dikaiakos for all the guidance, knowledge, opportunities, encouragement and - above all - the trust he put in me throughout these 8 years. I have learnt a great deal from our collaboration, both on an academic and personal level. Also I would to thank my research co-supervisor Associate Professor George Pallis for the advising, discussions and feedback he has provided me these years.

Second, I would like to acknowledge the support of my friends and colleagues at the Laboratory of Internet Computing (LInC), especially Andoena Balla and Maria Poveda who both have been here at the Department with me *"forever"* sharing my anxieties and frustrations from time to time. I also want to thank my colleaques Demetris Antoniades, Hariton Efstathiades, Christodoulos Efstathiades, Athanasios Foudoulis, Kyriakos Katsaris, Asterios Katsifodimos, Michele Mazzucco, Andreas Papadopoulos, Stalo Sofokleous, Antriani Stylianou, Nicolas Stylianides, Nicos Tsioutsias, Demetris Trihinas and all other present/past members of the lab, with whom I had extensive fruitful discussions and exchanged several inspiring ideas during the course of my studies.

Special acknowledgments go out to my academic collaborators Professor Athena Stassopoulou, Professor Liviu Iftode, Dr. Tamer Nadeem, Professor Harald Kornmayer, Dr. Harald Gjermundrod and Markus Knauer, for all the insightful com-

ments, motivation and knowledge they infused in me in order to widen my research and technical background.

Of course I cannot thank enough my closest friends Artemios Economides, Constantinos Anastasiou and Nikolas Karvellas for their constant motivation, engagement, laughs and great time we were having all these years together.

Finally, I would like to thank those truly extraordinary people on my side, without whom this PhD journey and its conclusion thereof would have never been possible. **Annita** my partner in life, who always backed me up, shared my success and failures and never stopped encouraging me to paddle ahead working harder and harder towards my goals. Last but not least, **Dad, Mom, my sister Efthymia, grandparents Kypros, Nicos, Eleni, Chrystalla and godfather Andreas** for believing in me, their untiring love, motivation and unending support (spiritual and physical). It is hard to put into words how much their support has meant to me.

This work was partially performed within the context of the project "V-Sense: A Vehicular Sensor Network Infrastructure" (PENEK/609/07), co-funded by the Republic of Cyprus and the European Regional Development Fund of the EU.

To my family…

Contents

List of Tables

List of Figures

Chapter 1

Introduction

During the last few decades, the availability of low-cost, high quality electronic circuitry, has realized the omni-presence of sensory equipment in relatively all aspects of our daily lives. Among the pioneers of this sensory ubiquitousness, automotive manufacturers, began fitting their vehicles with an extensive array of sensors that measure and record in high fidelity thousands of real-time vehicle-related (location, speed, acceleration, emissions, etc) and environment-related (temperature, humidity, CO2, etc) parameters. Their goal was and remains to be, the provision of safe, trouble-free and comfortable ride for vehicle drivers and passengers.

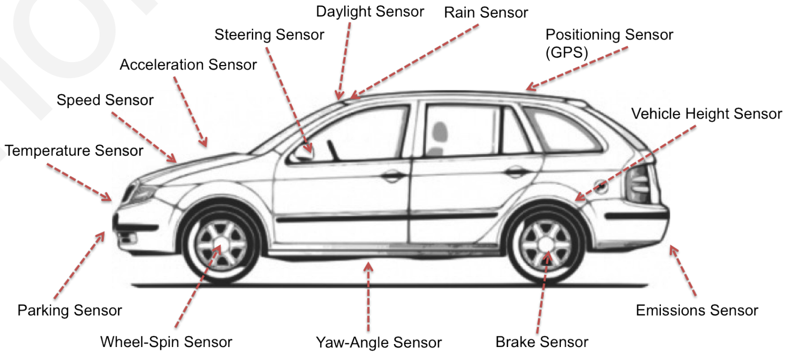


Figure 1.1: The Modern Vehicle

On a global scale, *Intelligent Transportation Systems (ITS)* aim to enable users and/or various transportation modes (including vehicles) to take informed decisions concerning commute safety and efficiency, by integrating real-time data and feedback from a number of heterogeneous resources. More recently, the concept of *Inter-Vehicle Communication (IVC)* has emerged and it is justifiably envisioned as one of the key supporting platforms for the establishment and market acceleration of ITS [31, 47, 48]. The allocation of the necessary wireless frequency band from governmental commissions around the world, provides concrete evidence that in the next few years, vehicles will feature wireless technologies as standard on-board equipment, supporting communication standards (IEEE 802.11p / WAVE [63]) designed exclusively for the vehicular environment. Acknowledging the potential societal impact of a "Connected Car", both the academia and automotive industry have for quite some time wholeheartedly supported the proliferation of international initiatives and consortia (PATH [16], Car2Car [1], CVIS [2], PReVENT [17]) to investigate this promising concept even further. It is expected that such technologies will ultimately facilitate the establishment of *Vehicular Ad Hoc Networks (VANETs)*, enabling thus the exchange and dissemination of a magnitude of sensory information among vehicles (V2V) and stationary road-side infrastructure (V2I).

Vehicular Sensor Networks (VSN) are a major departure from traditional mobile sensor paradigms, mainly due to the unique characteristics of the VANET and the capabilities of sensor nodes themselves [77]. In contrast to conventional battery-operated mobile sensors, vehicles have minimal energy constraints due to the existence of the engine. As a result, this source of ample energy enables them to have extended processing, storage and communication capabilities that are highly favorable while sensing and disseminating substantially larger quantities of information over wide geographic areas [114]. On the other hand, the high and variable mobility of the vehicles, as well as the presence of wireless signal obstructions, especially in urban environments (buildings, light poles, trees, etc), make IVC intermittent and hard to maintain or forecast in the course of the time [97].

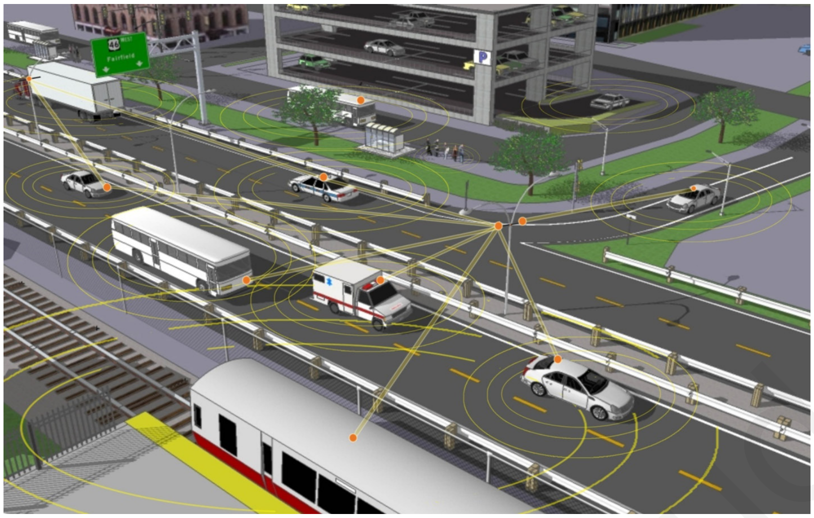


Figure 1.2: A Vehicular Sensor Network with V2X communication[1].

Nonetheless, the vision of such enabling communication and sensing technologies have paved the way for a number of interesting and innovative applications, which aim not only to combat and diminish safety-related road transportation problems but also to further improve the quality of peoples' daily commute. These application can be categorized to: safety message broadcast [121, 123, 127], congestion avoidance and resolution [46, 94, 119], content dissemination [85, 88, 95] and infotainment [36, 74, 113].

Traffic Information Systems (TIS) form another key category of non-safety VSN applications, wherein traffic information is sensed by vehicles and subsequently disseminated in near real-time to other vehicles through the help of V2V and V2I communication. TIS seek to enhance drivers awareness on the prevailing conditions on the road network, including traffic density and queues, road capacity and accidents. Their conceptual goal is to assist drivers to take informed decisions when planning their daily vehicular travel by avoiding existing or forecast traffic bottlenecks, ultimately resulting in the reduction of related socioeconomic and environmental impacts in modern societies.

---

[1] Image taken from U.S. DOT NHTSA

This PhD thesis focuses on presenting the scientific research performed in order to design and implement a *"Distributed Traffic Information System (TIS) based on V2X*[2] *communication for large-scale information diffusion in urban environments"*. To investigate the potentials and understand how such a distributed system will eventually be realized, we study the following problem domains: (i) **VANET Dynamics in Urban Environments**, (ii) **Traffic Sensing and Acquisition** mechanisms, and (iii) **Large-Scale Traffic Information Dissemination**.

More specifically, we explore the basic premise behind V2X, particularly the constantly evolving spatio-temporal nature of urban-based VANETs through which traffic information will flow. Here, V2X connectivity is modeled as sequences of contact graphs wherein complex network analysis measures are employed to gain insights on the structural properties of VANETs from a city-wide perspective. In contrast to other low-dimensional networks, we unveil the VANET's high sensitivity to fragmentation, extremely variable connectivity and total lack of small-world and scale-free features. Additionally, we present why and where particular network phenomena occur in the underlying road network and exploit the implications in the design of effective information dissemination mechanisms. Next, we consider the constrained bandwidth of VANETs, and thus we investigate the possibility of in-vehicle data caching. Particularly, we examine its capacity to improve the efficiency of TIS by reducing the time-to-serve requests and network overhead, while abiding established time and quality constraints. Through large-scale realistic simulative evaluations, we discuss that the use of simple TTL-based replacement policies can achieve significant improvements under both normal traffic conditions and unscheduled traffic events. Finally, we present V-Radar a reactive, VANET-based, traffic information dissemination protocol for metropolitan environments. In contrast to other approaches in the literature, V-Radar enables the querying and acquisition of traffic information along

---

[2] V2X sums up the communication among vehicles (V2V) and communication between vehicle-infrastructure (V2I).

a number of composite road-paths, starting from a vehicle's current position towards its final destination. Ultimately, this effective approach enables the driver to establish a more broad view of the traffic conditions that will be encountered further ahead in a timely manner. Finally, we present the analytical and simulation testbeds developed to support the above research, which are open-source released to the community for future studies.

The rest of this thesis is organized as follows. Chapter 2 presents the Motivation behind the research work performed in this PhD. Chapter 3 presents the System Model and Problem Statement for each of the three problem domains addressed in this thesis. Chapter 4 and 5 detail the approach and roadmap towards providing solutions for each problem domain. Chapter 6 outlines the related works. Chapter 7 presents the testbed established for evaluating the different research approaches and methodologies proposed for solving the above research problems. Chapter 8 presents the work on the V2X Evolutionary dynamics. Likewise, Chapter 9 presents the work on Urban Traffic Sensing and Acquisition, while Chapter 10 captures the research on Large-Scale Traffic Information Diffusion using Reactive Mechanisms. Chapter 11 provides conclusions and discussions on the work of this thesis, and finally Chapter 12 briefly describes Future Work plans.

Chapter 2

Motivation

Urban road traffic and congestion build-ups have an adverse effect in modern societies and the daily commute of citizens. Environmental pollution, millions of wasted money/hours and commuter stress are all the aftermath of vehicles stuck in traffic. In the United States alone, it is been estimated that the cost of traffic congestion for 2012, exceeded $121 billion dollars, while the yearly amount of delay endured by the average commuter was approximately 38 hours [84]. On a global scale, the delay estimations are startling and alarming, with vehicle users wasting approximately 90 billion hours in traffic jams each year.[1] As these numbers are expected to rise substantially over the next years and given the global economic recession, it is crucial both for the research community and the automobile industry to identify and provide the necessary solutions in an effort to improve daily road transport and minimize unnecessary expenditure.

Vehicular traffic estimation and prediction systems currently in operation in metropolitan areas around the world, require the use of dedicated roadside infrastructure (i.e, magnetic loop detectors, piezoelectric sensors, cameras). In addition to the extremely high deployment and maintenance costs, such systems and their

---

[1] The Economist (2014) - *Connected Cars will make Driving Safer, Cleaner and more Efficient. Their introduction should be speeded up*, September 2014

components are usually accompanied with low reliability, limited coverage and short life expectancy [54]. Considering the expected rise in the number of vehicles in the near future, as well as the sophistication and high reliability of modern electronic circuitry, one can clearly see that such systems will eventually be rendered obsolete. Newer solutions based on Floating Car Data (FCD) monitoring, i.e information about GPS localization of vehicles, provide interesting alternatives, however they inherently require the offloading of massive amounts of information to central Internet based servers via the already saturated public - but privately regulated and monetized - cellular network [30].

VANET-based *Traffic Information Systems (TIS)* hold a prominent role in the long-term strategy of mitigating the aforementioned urban traffic implications. Their primary function is to facilitate dynamic route planning through "almost real-time" collective traffic measurements [46, 53, 71, 107] obtained solely from VSN-capable vehicles. TIS have been envisioned to provide drivers and/or in-vehicle navigation systems with a more broad and fine-grained view of the traffic conditions that will be encountered in-between two locations. The key question that needs to be systematically and accurately answered by such systems is: *"What are the prevailing traffic conditions en-route to a particular destination?"* - ultimately, such knowledge can be extremely valuable in the process of calculating more optimal route(s) by taking under consideration various constraints such as total travel time, fuel cost and emissions. In addition, collective road traffic estimates reported by vehicles can be used by transport authorities for identifying: *"What is the localized traffic behavior of different regions in the city?"* - such knowledge can assist authorities to implement and optimize dynamic traffic signalization, adaptation of road segment(s) speed limits, on-ramp metering, greenhouse gas (GHC) levels and even the enforcement of driving laws.

This PhD thesis is motivated by the above societal problem, which is major in all over the world during the last-half of the 21^{st} century. In line with European Commission's Action Plan[2] on ITS (2008) and Directive 2010/40/EU (2010) [100], it aims at contributing and accelerating the vision of congestion-free, sustainable and economic mobility, by proposing a *"Distributed Traffic Information System (TIS) based on V2X communication for large-scale information diffusion in urban environments"*. However, if one considers the intrinsic properties of VANETs, the realization of such a large-scale V2X distributed system, especially for metropolitan areas, becomes a non-trivial endeavour. In contrast to the highway setting, vehicular traffic in urban environments is anything else but uniform and reliable. On a macroscopic scale, the societal activities of people during diurnal cycles, along the topological structure and features of the underlying road network (including traffic light signalization), cause city-wide car flows to be non-linear and less predictable. Zooming inwards to the microscopic scale, individual driver behavior, vehicle technical characteristics (acceleration/deceleration rates, length, wheelbase, etc.) and unexpected events (accidents, mechanical breakdowns, etc.) contribute even further to the non-uniformity of urban mobility distribution. All these factors greatly affect the following 3 supporting pillars, required for effective and efficient information diffusion in VANET-based TIS. Each of these pillars can be seen as a different problem domain of its own that justifies for a comprehensive investigation:

- **Network Dynamics** - mobility is key in the establishment of peer-to-peer interconnection among vehicles. It dictates *when, where and for how long*, ad hoc links can be formed, which in turn facilitate any potential information exchange among neighboring vehicles. These dynamics - both for direct and multi-hop links - along with the spatial density and penetration ratio of V2X-capable vehicles, are responsible for the oscillation in the instantaneous network

---

[2] `http://ec.europa.eu/transport/themes/its/road/action_plan/`

topology, specifically its state transition between well-connected and sparsely connected [93, 116]. In the case of the latter, frequent abrupt changes in the distribution and density of ad hoc links in the topology, can push the network to fragmentation for a prolonged period of time, a phenomenon that naturally impedes the end-to-end communication capabilities among vehicles located in distant geographic areas. As a result, large-scale traffic information diffusion becomes a non-trivial task.

- **Urban Traffic Sensing and Acquisition** - each of the road type (i.e urban, collector, arterial, freeway) a vehicle travels on, presents a different set of opportunities and requirements as far as traffic information sensing is concerned. This is due to the inherent mobility patterns that prevail on each road type (density and distribution of vehicles, speed limits, number of lanes, presence or absence of intersections, etc.). Consequently, accurate calculation or prediction of vehicle traffic (location, speed, acceleration, etc.) in the underlying street topology in a distributed manner, requires network and location-aware adaptive processes that should involve, as much as possible, the collaborative efforts of VANET-enabled vehicles. Although, each individual vehicle can report its own localized view of its surrounding neighborhood through through promiscuous sniffing of the wireless channel, it is also possible to provide an approximate view of an extended geographic region by collaborating with other vehicles it encounters during of its travel through the city. Given the dynamics of the VANET topology and the limited resources of the channel, each vehicle should employ mechanisms that allow its prompt reaction to any events regarding connectivity status changes. Therefore, information freshness and response time are kept within acceptable margins, while avoiding additional network overhead.

- **Message Dissemination** - spatio-temporal dynamics and characteristics of VANET topologies, render message dissemination mechanisms for traditional mobile ad hoc networks (MANETs) incapable for efficient traffic information diffusion in large scale environments. An amalgamation of wireless broadcast, geocast, multi-hop, carry-and-forward and hovering communication paradigms are required [106] each satisfying a particular network property. In the case of carry-and-forward paradigm, such schemes should capitalize their success on cache-replacement algorithms that consider the specifics of the urban vehicular environment and the traffic information disseminated within.

Chapter 3

Problem Statement

This section presents the overall System Model, the research objective to be met and the specifics of the problems domains which will be researched in the context of this PhD thesis.

3.1 System Model

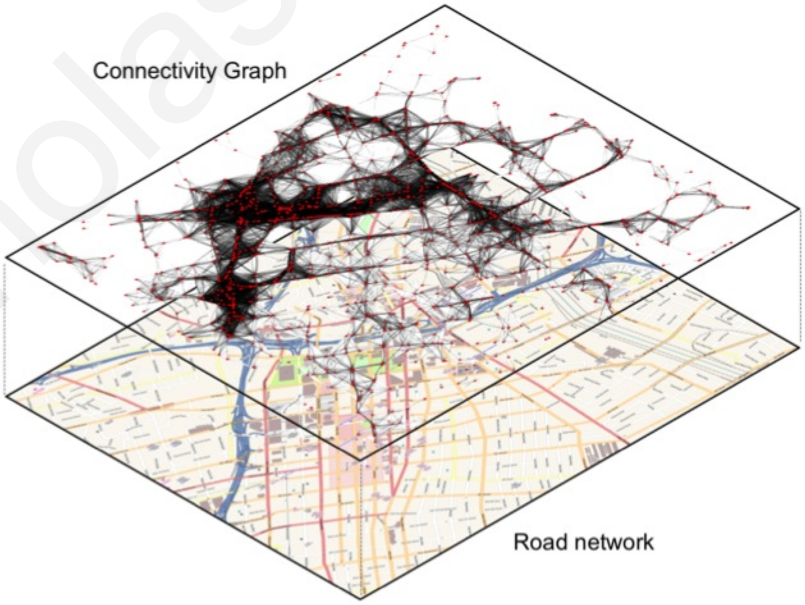


Figure 3.1: The VANET Connectivity Graph

We consider a Vehicular Ad Hoc Network (VANET) in a large-scale urban environment, comprised of modern vehicles equipped with an IEEE 801.11p capable transceiver. All vehicles share an identical nominal line-of-sight and non-line-of-sight transmission range (R_{LOS} and, R_{NLOS} respectively). Each vehicle has its own mobility profile defined as an ordered-sequence of spatio-temporal points $MP = \langle P_1 \ldots P_n \rangle$, where $P_c = (x, y, t)$ and x, y are defined as geospatial coordinates (transformed to the Euclidean space) and t is a distinct time instance. $S = t(P_1)$ and $D = t(P_n)$ are regarded as the start and destination location/times, respectively, while $T_{travel} = t(P_n) - t(P_1)$ being the total travel time per vehicle. Vehicles indicate their intention to participate in the VANET by frequently broadcasting beacon messages B_{msg} of the form $(V_{id}, P_c, \upsilon, H, R_{id})$, where V_{id} is the vehicle unique ID and (P_c, υ, H) are respectively its current position, speed (m/s) and heading (°). R_{uid} denotes the unique ID of the road the vehicle is currently travelling on. Information regarding the underlying road-network topology is provided to the vehicle through on-board preloaded digital maps, while its current geographic location is obtained via the Global Positioning System (GPS). Maps can also be enriched with historical statistics that exhibit the traffic conditions of the road-network at different times of the day.

On one hand, vehicular commuters would like to be kept updated on the road traffic conditions between S and D in order to optimize MP towards the best possible T_{travel}, given all safety and legal considerations. On the other hand, transport and planning authorities wish to have a near real-time view of traffic flows and events in various parts of the city so as to alleviate the possibility of any future critical situation, or act upon one if necessary. To this end, vehicles constantly sense the traffic density and average speeds in their surroundings by promiscuously listening the wireless channel for other vehicles' B_{msg}. Traffic reports (I) can materialize as singular information items $I_{(VID)}$, i.e report from just vehicle with V_{id}, or aggregate information items $I_{(ROI)}$ from many vehicles that reside within a particular region of interest (ROI). Without loss of generality, we model a ROI as a circle with center $C = (x, y)$ and radius R_r, which can possibly contain both high density and empty road segments.

Each vehicle encompasses a local cache where it temporarily maintains traffic reports sensed from the roads that have been traversed. Regardless of the fact that TIS fall in the category of Delay Tolerant Network (DTN)-like systems, cache-revalidation mechanisms are utilized to guarantee information freshness and minimize response rates. Depending on the network circumstances, reports can be diffused in the road network using an augmentation of broadcast, geocast, multi-hop, carry-and-forward and hovering communication paradigms. Vehicles learn about the traffic conditions between S and D by reactively issuing traffic queries towards a particular ROI or a number of road paths. Any road path P is a defined as a unique, alternating sequence of connected intersections and roads. Finally, traffic information about a ROI can be proactively prefetched to other regions that have expressed or shown to have an interest.

3.2 Objective

"To design and implement a robust and efficient Traffic Information System, based on Vehicular Ad Hoc Network communication, adaptive to network dynamics and constraints, that reports a close-to-accurate representation of spatio-temporal urban traffic conditions over a wide geographic area."

Notation	**Description**
V_{id}	Vehicle UID
P_c	The current geospatial coordinates of a vehicle
MP	The mobility profile of a vehicle
S, D	Start and destination location of a vehicle
T_{travel}	Total travel time per vehicle
B_{msg}	Beacon message
u	Vehicle Speed (m/s)
H	Vehicle Heading (°)
R_{uid}	Unique ID of a road
ROI, R_r, C	Circular Region of Interest, center coordinate C and radius R
$I_{(VID)}$	Traffic report from a single vehicle
$I_{(ROI)}$	Aggregated traffic from several vehicles in a ROI
L_{ij}	Direct link between any vehicles i and j
LD_{ij}	Link Duration

Table 3.1: Notation used in the System Model and Problem Statement

3.3 Problem Domains

3.3.1 VANET Dynamics in Urban Environments

As a result of the characteristics of urban vehicular mobility, VANETs that will eventually be established over such environments will be highly dynamic. This makes them susceptible to various factors that have the potential to degrade the desired Quality of Service (QoS), or at the very worst invoke prolonged network instability and disruption.

In this respect, when discussing about *a highly dynamic* VANET, we are referring to an ad hoc network topology, which changes very frequently. Firstly, this is due to vehicles (nodes) that can abruptly joir or depart the communication plane, thus instantly making themselves available or unavailable for information exchange. Secondly, buildings, light poles and general obstructions cause what is called as the "urban canyon" phenomenon, that has negative side effects in the propagation and strength of wireless signals. As consequence of the above, communication links in this system are very volatile and transient. The probability of two neighboring vehicles I

and J establishing a direct communication link $L_{ij}(t)$ among them, is a function of MP and R. That is, I and J will be able to communicate directly (1-hop), if both are within transmission range of each other for a continuous time duration that is $LD_{ij} \leq T_{travel}$. However, the duration and strength of L_{ij} is influenced by H and υ of each communicating party. Adding to the above, the probability of two remote vehicles exchanging information over multiple hops, is a function of the collective MP of all intermediary vehicles in the particular communication path. As a result, multi-hop communication is considered to be even more fragile and unpredictable than direct communication.

Consequently, the following networking factors pose problems to information exchange over large-scale urban-based VANETs and deem for methodical observation and understanding:

- **Ad Hoc Link (1-hop) Characteristics -** (i) *What is the average duration of 1-hop links?* (ii) *How often does a distinct vehicle-pair establish a direct link?* (iii) *How long does it take to re-establish a broken direct link?* and (iv) *What is the temporal and spatial distriubution of 1-hop links in the road topology?*. Answering these questions is a significant step in delineating the opportunistic nature of ad hoc links among vehicles and understanding their spatio-temporal formation and lifetime features. By obtaining insights on 1-hop link characteristics, we can make approximations on when and where TIS information can be exchanged among neighbouring vehicular peers, including restrictions on packet size and frequency of information updates.

- **Fragmentation -** (i) *How often does the VANET gets fragmented?* (ii) *What are the factors that invoke partitioning or fusion of the network core and its peripheral components?* (iii) *What is the spatial distribution of fragmented components and how large (in terms of population and area) can they become?* (iv) *Are fragmented components connected to particular geographical features?* (v)

What is the internal connectivity levels of fragmented components?. Answers to these questions may capture the dynamic capabilities of fragmented components and even possibly the existence of dense communities established inside a VANET that would eventually sustain TIS information diffusion. Primarily, we will be able to understand whether (and which) network fragments are connected sufficiently to support TIS, and what factors drive such strong attachments. Secondly, we can examine the extend to which network partitioning introduces barriers in large-scale information dissemination, both in terms of time and geography.

- **Multi-Hop Communication Path Characteristics -** (i) *What size can multi-hop paths attain?* (ii) *What is the maximum geographic distance between two remote vehicles connected via multi-hop communication?* (iii) *What is the temporal and spatial distribution of multi-hop paths?* (iv) *What is the average lifetime of a multi-hop path and what are the factors that cause it to break?* (v) *Is there path redundancy among any two vehicles?*. Answering these questions ultimately allows us to estimate the feasibility and maximum potential of traffic information diffusion in the case of a multi-hop, path-sufficient, large-scale urban VANET and thus the geographic limitations that TIS dissemination mechanisms must work within. Moreover, they disclose the network properties that should be monitored and considered in order to improve multi-hop reliability.

- **Node Importance -** (i) *How does the concept of 'importance' can be abstracted for the VANET environment?* (ii) *Are there any vehicles whose position in the VANET topology permits them to influence global information diffusion?* (iii) *How many 'important' vehicles can be identified and can their mobility profiles be modelled?* (iv) *What is their spatial distribution on the road topology?* (v) *What is the temporal evolution profile of node importance?*. Answering these questions enables us to pinpoint the features of so called 'important' nodes, which in the end might be strong candidates for uptaking key roles during TIS data diffusion processes. Decisions for TIS information exchange can consider the mobility profile of such nodes and subsequently adapt their delegation mechanisms in a manner that utilizes them as frequently as possible so as to achieve acceptable levels of QoS.

3.3.2 Traffic Sensing and Acquisition

Core to TIS architectures are the mechanisms that embody accurate sensing of the prevailing vehicular traffic status in the underlying road topology. These require distributed, collaborative and adaptive procedures that take under consideration not only the aforementioned network dynamics, but also the road type each vehicle is driven on and its individual MP. Each vehicle is expected to make consistent and acceptable approximations in terms of the traffic sampling method and frequency, as well as the particulars of temporal storage. The goal is to make the set of all localized I_{VID} and aggregated I_{ROI} traffic details in its disposal (sensed both by itself and others), readily available for consumption by its road peers, without imposing additional overhead to the already limited network resources.

- **Localized Sensing Details -** (i) *What is the correct sampling frequency of traffic information?* (ii) *Given a particular MP, where are the best locations in the road network to sample?* (iii) *What are the semantics of a distributed traffic estimation?* Given that continuous traffic recording (even by a small fraction of V2X-enabled vehicles in a city) will eventually result to immense amounts of generated data, that neither will be wise to store nor disseminate. Answering these questions will disclose details on how to perform efficient and effective traffic estimation in a distributed ad hoc manner through the use of partially but intelligently sampled data.

- **Aggregate Sensing Details -** (i) *How can aggregate traffic reports can be encoded?* (ii) *When and where aggregation procedures need to be triggered?* (iii) *How and which extend information accuracy is maintained in such aggregated reports?* Answering these questions allows us to establish the laws that will govern traffic information aggregation schemes. In this context, we will be able to observe where and how aggregates can be encoded without any substantial loss in accuracy when representing the prevailing traffic situation within a confined geographic region of interest. In addition, we will be able to determine and evaluate those circumstances where triggering aggregation actions can yield improved city-wide traffic awareness without sacrificing network utilization.

- **Caching -** (i) *Is caching a viable solution for* $I_{(VID)}$ *and* $I_{(ROI)}$*, in order to overcome network fragmentation?* (ii) *What is the lifetime of cached information?* (iii) *How is information accuracy affected when using different cache replacement schemes?* Answers to these questions will exhibit the feasibility of caching mechanisms for highly dynamics environments such as the urban-based VANET. Although caching is proven to work in traditional mobile ad hoc networks, it might not result in the same advantages when dealing with highly variable traffic information in the context of TIS. In addition, such questions allows us to examine how information accuracy behaves in the presence of caching, both for individual and aggregated traffic reports.

3.3.3 Large-Scale Information Dissemination

Distributed traffic reports ($I_{(VID)}$ and $I_{(ROI)}$) by VANET peers will provide vehicles in a metropolitan area with a view of the prevailing traffic conditions en route to their destination. Moreover, they enable authorities to establish traffic maps describing flows and unexpected events city-wide. For these to be possible, sensory information needs to be disseminated (proactively or reactively) over a large-scale geographic area with dynamic network topology. At first, MANET forwarding protocols are unsuitable for such topologies, thus VANET-specific protocols are needed. However, their performance in the complex urban environments we are investigating is not yet proven. Secondly, reactive dissemination can be achieved using multi-hop query-reply methods, however these might encounter difficulties due to fragile lifetime of such connectivity paths. To this concern, proactive information prefetching to suitable locations can possibly constitute part of a hybrid solution.

- **VANET Routing Protocols -** (i) *Which VANET-specific routing protocols are adequate for TIS exchange?* (ii) *How these perform in large-scale urban environments under representative mobility and system conditions?* (iii) *How such protocols can be improved to maintain TIS QoS requirements?* A large body of work related to VANET routing protocols is available in the literature, but only a small number of the proposed solutions have been tested under those demanding mobility, network and application-oriented conditions for which they were envisioned. Answering these questions allows us to evaluate the extend to which such solutions are capable to support TIS exchange within the expected QoS margins, and what improvements in their architectures might be necessary to achieve them, thereof.

- **Reactive Traffic Information Queries -** (i) *How can $I_{(VID)}$ and $I_{(ROI)}$ be reactively obtained using a query-reply model?* (ii) *What are the semantics of traffic query language and architecture?* (iii) *What are the system model parameters that guide an adaptive query-reply system?* (iv) *How does a reactive query-reply model performs (accuracy and response time) over the available communication paradigms?* Although routing protocols are key in dissemination processes, they require traffic information first to be readily available prior forwarding. Answering these questions allows first to examine how to accomplish reactive TIS dissemination mechanisms in large-scale urban environments.

Such mechanisms rely on query-reply paradigms, where a source vehicle poses geo-tagged traffic queries, and other vehicles provide a specific reply, either solely or collaboratively. In addition, through careful evaluation we are presented with quantitative and qualitative insights on the performance of such mechanisms.

- **Proactive Traffic Information Prefetching -** (i) *Can a query-reply model be enhanced with proactive traffic information prefetching?* (ii) *What are the system model parameters that guide prefetching?* (iii) *When and where information should be prefetched?* (iv) *How does prefetching performs (accuracy and response time) over the available communication paradigms?* Given the nature of VANET communication and existence of fragmented components, reactive query-reply models based on multi-hop communication may result in sub-par performance. In the context of TIS, prefetching allows specific traffic information to be readily available on location or in the proximity of certain vehicles that might be require it in the near future. Thus it removes the necessity of frequently posing multi-hop queries towards a particular ROI. Answering these questions we can examine the conditions that drive prefetching, which factors drive and invoke such actions and evaluate their added-value in hybrid dissemination model performance.

Chapter 4

Approach and Methodology

This section presents the approach and methodologies that will be taken in order to provide solutions to the problem presented earlier in Section 3.

4.1 VANET Evolutionary Dynamics through Complex Network Science

We decide to study the spatio-temporal evolution of VANET communication in urban environments from the aspect of *Complex Network Science* in order to observe and derive crucial insights in their extremely dynamic nature. The motivation behind this decision stems from the fact that the study of the structural properties of large, real-world, dynamic systems (i.e. the Internet topology, collaboration, biological and on-line social networks) has lead to crucial observations having significant influence in information and social sciences. For instance, the discovery of power-laws in the Internet topology by M.Faloutsos et al.[49], enabled the design and implementation of efficient information routing protocols that capture the Internet topology characteristics.

Methodology - To achieve the above, we need both simulative and extensive analytical processes. Particularly, we model the VANET as an undirected graph, where vehicles correspond to the set of vertices and communication links correspond to the set of edges. Having in mind the System Model presented in Section 3 while data mining a large-scale highly realistic mobility trace, allows us to study a hypothetical instantaneous ad hoc network that could potentially emerge given vehicles' travel trajectories and proximity interactions thereof. Hundreds of such time-sequenced undirected graph snapshots are analyzed using a large body of complex network science algorithms to capture interesting insights on the evolutionary dynamics of VANETs. The aim of this is to provide a "higher order" knowledge of the time-evolving topological characteristics of the VANET communication graph, as compared to the "first-order" knowledge provided by the studies in the existing literature, and also to extend to examine the VANET dynamics from a more specific spatial viewpoint in order to identify where/why in the underlying road network, particular network phenomena occur. Furthermore, we make reference and correlate our findings with the main communication paradigms employed in VANETs, and discuss whether the emerging network phenomena can indeed support (and to what extend) these paradigms.

4.2 Urban Traffic Sensing and Acquisition

We approach solutions to the particular problem domain through the modelling of different urban traffic sensing and acquisition procedures as distinct components of an optimization problem that targets to maximize the overall throughput of a VANET-based TIS. Moreover, it is important that the optimization problem at-hand should examine how acceptable levels of information accuracy can be maintained, while reducing unnecessary network overhead. To this extend, we consider this as a two-side approach, with one side involving the identification of those parameters that drive efficient traffic sensing, both for localized and aggregated views. The second part evolves around the behavior and applicability of caching mechanisms for minimizing network overhead and TIS response times. Proposed schemes that address the above maximization problem will eventually be encapsulated in well-defined protocols that in turn can be quantitatively and qualitatively evaluated through extensive simulations.

Methodology - The task of accurately sensing vehicular traffic is dependant on the underlying VANET connectivity conditions, as well as the road type each vehicle is driven on in accordance to individual mobility profile. By considering the findings of the VANET evolutionary dynamics both in time and space, we will design schemes for efficient and effective traffic sampling. In terms of caching, we examine its applicability for dynamic information that flows over VANETs and its comparative performance against non-cache schemes. Particularly, our work examines the implications of caching on the efficiency of VANETs and on the quality/accuracy of TIS. Our study is performed in the context of VITP [43], a proactive, location-oriented protocol designed for the retrieval of dynamic vehicular information over V2X communication. Again, under realistic mobility scenarios, we investigate the capacity of caching schemes and whether they can improve further the efficiency of VITP by reducing the time to serve information requests, minimizing overall network resource consumption, and improving information accessibility in the presence of mobility constraints that result to short-lived links, network disconnections, etc. We conduct an extensive exploration of the trade-off between the efficiency and the quality of VANET-based TIS services, in the presence of caching.

4.3 Large-Scale Traffic Information Diffusion using Reactive and Proactive Mechanisms

We decide to examine how traffic information can be diffused in a large-scale urban environment, either reactively or proactively. However, this decision requires the presence of adequate and capable underlying routing protocols. We initiate our quest by evaluating the performance of recently proposed (and highly cited) VANET-dedicated routing protocols in large-scale urban environments, under realistic mobility and traffic application constraints. The motivation behind this decision is the fact that such protocols are constantly evaluated under naively simple conditions (i.e. mobility, road topology) and in the absence of applications that would impose system and network loads similar to the ones of a real deployment. Such conditions could report highly biased performance results in a manner that over/under-estimate their capabilities during traffic information dissemination. To this end, we implement a large-scale simulation testbed for realistic VANET protocol evaluation. In this context, we also provide reference implementations (from scratch) for a number of VANET routing protocols, each one supporting a particular communication paradigm. Our final aim is to provide the complete test-bed available to the research community through a well-established open-source network simulation framework.

Methodology - As a first step we target how to address reactive traffic information dissemination. Particularly, proposing a novel vehicular traffic information query protocol for urban environments based on V2V communication. In contrast to other approaches in the literature that aim to obtain traffic information on singular roads, the proposed protocol seeks to enable the querying and acquisition of traffic information along a composite road-path, starting from a vehicle's current position towards its final destination. Specifically, the protocol should be able to query not only the initially selected road-path, but also a number of alternate paths that lead to the vehicle's destination. Eventually this will allow the driver or the in-vehicle navigation system to establish a more broad and complete view of the traffic conditions that will be encountered further ahead. Furthermore, we will examine how the protocol can adaptively change its core functions (query rate, available bandwidth, etc) based on the ever changing network and mobility constraints. In the second part, we examine whether proactive traffic dissemination, namely information prefetching, can indeed provide added-value to a multi-hop query-reply scheme such as the above. Specifically, we examine the system model parameters that guide/invoke prefetching actions and thus propose an adaptive mechanism that realizes the aforementioned.

Chapter 5

Thesis Roadmap

This sections provides an overview of the conceptual roadmap that was followed in order to meet the goals and objective of this PhD thesis.

1. **Related Work:** This entailed the trawling of the VANET bibliography to identify and collect research works related in any manner to the three problem domains presented in the PhD thesis. Following standard practices, collected articles were categorized, tagged and read in order to gain the necessary knowledge for future reference and consideration. Mendeley [13] was used for cataloguing all related work. This was an ongoing and repetitive effort throughout the duration of this PhD.

2. **Mobility DataSet:** This milestone involves the collection or generation of real and realistic vehicular mobility traces, respectively, that were utilized in the numerous simulative exercises throughout the duration of this PhD. A number of small, publicly available real traces (from taxis, buses and private vehicles) are examined to verify their temporal granularity and correctness, with different spatial extrapolation techniques employed to complete them. On the other hand, traffic modelling [99], microscopic traffic simulators [21] and conversion

tools [25] are utilized to generate synthetic mobility traces guided by the laws of well-established mobility models on real road topologies.

3. **Simulation Testbed:** A VANET-dedicated simulation test-bed has been developed over the well-known ns-3 [23] framework. In the context of this milestone, a number of VANET routing protocols were implemented from scratch so as to facilitate the evaluation of proposed information sensing, diffusion and storage schemes in the future. The necessary tooling was created in order to allow the use of real world maps [14] during simulation scenarios. The aim was to stress-test all components in a scenario (routing protocols, applications, network) so as to resemble as closely as possible the real and demanding conditions for which they are envisioned. Optimizations to the test-bed code based were made in order to facilitate large-scale (spatial and number of nodes) and lengthy experimentation scenarios.

4. **Reference Applications:** Albeit brief, this milestone is very significant, since it contains the proposal and finalization of reference applications that will drive the novel TIS concepts and techniques proposed in this PhD thesis. Particularly, two vehicular traffic applications and their architectures were defined. The first is a reactive traffic information protocol for urban environments based on V2V communication, that meets the performance requirements of TIS systems [114]. This applications should be able to acquire the prevailing traffic conditions in a composite road path between two geographic coordinates - usually start/destination - using multi-hop query reply schemes. The second reference application dealt with providing traffic maps of large metropolitan regions to transport authorities. Such a system should employ a hybrid scheme with reactive and proactive traffic acquisition methods. For the latter method prefetching information techniques were employed.

5. **VANET Evolutionary Dynamics:** Here, the spatio-temporal evolution of large-scale urban VANETs are analyzed, following complex network science techniques. This step includes the implementation of the necessary programmatic tools to transform the aforementioned mobility datasets into undirected graphs that model the instantaneous wireless connectivity of vehicles. For graph generation, the tools should consider important system model parameters such as communication technologies and vehicle penetration ratio. Most importantly the tool must be able to run an array of social network analysis (clustering, centrality, community detection, distribution fitting, etc) algorithms and statistics on the generated graphs and provide explanatory figures. Off-the-shelf software libraries were used whenever available. The results are examined in depth and interpreted accordingly in order to derive meaning full answers to the research questions posted earlier in Section 3.

6. **Traffic Sensing Algorithms:** In the context of this task, we propose and evaluate novel mechanisms that embody accurate sensing of the prevailing traffic status in the underlying road topology. Particularly, we investigate different sensing parameters such as sampling frequency and optimal sampling geolocations and discuss the semantics of distributed traffic estimation. Given that for the purposes of traffic sensing we wish to examine information aggregation concepts, we propose encoding schemes influenced from MANET paradigms and subsequently evaluate their effect on information accuracy.

7. **Traffic Information Caching:** Here we evaluate the feasibility of caching mechanisms for highly mobile environments such as the urban-based VANET. Considering dynamic information encapsulating the prevailing traffic status, different cache-replacement schemes will be tested in order to disclose important functional requirements of carry-and-forward protocols.

Chapter 6

Related Work

This section provides an overview of various research works which were recorded in the literature during the past years and are related to the research direction uptaken to meet the goal of this PhD thesis.

6.1 Vehicular Sensor Network Infrastructures

Various research works were proposed in the literature during the past years and are related to the design and establishment of Vehicular Sensor Network (VSN) infrastructures.

CarTel [62] initially advocated the concept of a distributed mobile sensor computing system, wherein vehicles collect, process and disseminate large amounts of heterogeneous sensory information (automotive diagnostics, pollution, geo-tagged images and video, etc) while being driven in an urban environment. Specifically, vehicles utilize their on-board sensing hardware to retrieve and analyze data locally, before sending them via a communication infrastructure to a central storage repository for further analysis and visualization. To this end, the CarTel system is composed of three core components: (i) a Web-portal, (ii) an intermittently connected database called ICEDB and (iii) CafNet: a delay-tolerant network stack. The Web-portal serves as the interface of the system with the users and also as the central repository where various applications are hosted. Through the ICEDB continuous query interface, applications issue SQL-like queries to vehicles, specifying what sensor data and at what rate are required. Depending on the application type and consequently the amount of sensory information that eventually will be generated and disseminated by vehicles (i.e. sensor nodes), users can additionally define filtering, aggregation and prioritization policies, in order to facilitate the prompt delivery of data back to the portal. In the CarTel model, the Web-portal also uptakes the role of a traditional "sink". New or updated versions of continuous queries are pushed to a vehicle, as soon as it connects to the portal via opportunistic wireless connectivity (i.e. open Wi-Fi access points). Due to the inherent intermittent connectivity of a wireless network, the CarTel system widely employs on the vehicles the CafNet network stack. CafNet, is a general-purpose message oriented network stack that departs from the traditional data sending paradigm, where an application injects messages in the network as soon as it has something to send or utilizes a FIFO buffering schemes in the absence of network connectivity. On the contrary, CafNet's goal is to instantly notify applications when connectivity over one or more communication mediums (Wi-Fi, Bluetooth, etc) becomes available or changes status. This allows the applications to decide what data should be transmitted and with what priority. At the same time, if there are any pending results in the vehicle from previous continuous queries, these are streamed to a relational database at the portal.

Given however the large amount of data that modern, high-fidelity, in-vehicle sensors can generate and also the lack of sufficient a priori knowledge for imposing effective filtering mechanisms, the conventional approach of transmitting the data to a "sink" for further analysis, is not practical in VSNs. In contrast to CarTel, the MobEyes [75, 77] system exploits IVC-enabled mobile collector agents instead of relying on the wired Internet infrastructure, thus improving the overall robustness and scalability. Specifically, in MobEyes, sensed data are cached locally at the vehicle. Various sensing equipment can be fitted on the vehicle, therefore a Mobile Sensor Interface (MSI) enables uniform data access and retrieval. The processing capabilities of the node itself are utilized for extracting various features of interest. Using the MobEyes Data Processor (MDP), each participating vehicle periodically generates short summaries of the extracted features and by exploiting its mobility, it disseminates them to other vehicles in an opportunistic fashion.

Packets containing one or more generated summaries have a fixed header format that includes the packet type, the generator ID, a local unique sequence number, timestamp and the generators current geo-coordinates. Since summaries are time and location sensitive, each summary is additionally attributed with a timestamp and location metadata. The MobEyes Diffusion/Harvesting Processor (MDHP) is responsible for the dissemination of the summary packets to other vehicles. MDHP supports two diffusion strategies: (i) single-hop passive diffusion, where each packet is allowed to reach only the one-hop neighbors, (ii) k-hop passive diffusion, where each packet is allowed to travel up to k-hops from the source node. Based on the received summaries and accompanied metadata, harvesting agents can identify data of interest and consequently contact the source vehicle to extract the whole sensory dataset. In conjunction with passive diffusion, harvesting agents can proactively query neighbouring vehicles so as to collect missing summaries for a particular geographical area. Thus, MobEyes exploits the concept of Bloom filters [33], a space-efficient probabilistic data structure for membership checking. The harvesting agent uses a Bloom filter to represent already acquired summaries, utilizing the generator ID and unique sequence number in the summary found in the packet header, and attaches it to a transmitted “harvest” request message. Neighbors check their local cache and in the case that one or more missing packets are found, these are returned to the harvester piggybacked to an acknowledgment message. In turn, the ability to exchange Bloom filters among multiple harvesting agents creates a distributed, partially replicated sensory information index.

In [68], the authors propose a hybrid (vehicular and static node) VSN framework to monitor road traffic conditions and provide desired and reliable information for users, particularly for vehicle drivers that want to plan their travelling trips. On one hand, due to their mobility in a urban environment, vehicles can acquire more fresh information, that is spreading on the map out of the reach of static sensors. On the other hand, static sensors can gather and store information from passing-by vehicles. Road-side sensors are placed at the beginning and end of road-segments, which in turn are naturally delimited by signalized intersections. These sensors are equipped with a local database that stores self-observed traffic information on the given road-segment. In addition, they are capable of processing and storing information sensed by passing-by vehicles. Upon entering a road, vehicles query the road-side sensor on the prevailing traffic conditions of the segment and their destination. The sensors reply back with the required traffic information, while the vehicle uploads recorded traffic information about other segments traversed earlier in its trip. Several vehicles can form an ad hoc group if they are in each other's communication range and moving in the same direction. Vehicles entering the group exchange their local and group information with their 1-hop neighbors. Maintenance messages are periodically broadcasted to maintain the group structure. As time progresses, newly sensed information are shared internally among vehicles and eventually all group members posses the same data. However, installing and maintaining a huge number of fixed road-side sensing equipment, imposes an excessive overhead both in terms of money and human effort. In [42, 43], the authors propose the Vehicular Information Transfer Protocol (VITP), a proactive, location-oriented protocol designed for the retrieval of dynamic vehicle-sensed information over VANETs. VITP enables the formation of an on-demand, location-aware, virtual ad hoc server (VAHS) for the resolution of queries that would otherwise be served only through road-side units. For instance, in contrast to the approach taken in [68], where the prevailing traffic conditions are obtained by querying the nearest road-side unit, VITP empowers a group of neighbouring vehicles (VITP peers) to form a VAHS and contribute to a road-traffic information query. Specifically, VITP allows vehicles participating in a VAHS, to summarize location sensitive information (i.e. average speed in a road-segment) and transmit them back to the query originator. VITP specifies the syntax and semantics of messages (queries/replies) exchanged between the peers in the infrastructure. A VITP transaction consists of four phases: (i) Dispatch-query phase: a request Q is transported through the underlying VANET toward its target area L. Q goes through a number of intermediary VANET nodes, which push the message toward its destination using geographic routing. (ii) Virtual Ad Hoc Server (VAHS)-computation phase: the VITP request is routed between the VITP peers of the VAHS. VAHS consists of the VITP peers that contribute to computation of the reply to a VITP query. (iii) Dispatch-reply phase: the VITP reply is geographically routed toward source region. (iv) Reply-delivery phase: broadcasts the VITP reply to the VANET nodes of source region, so that the reply can be received by the VITP peer that originated the transaction. Proactive protocols in the likes of VITP can avoid network saturation caused by flooding and diffusion strategies since information is queried on demand rather than being pushed periodically during discrete time intervals.

6.2 Dedicated Routing Protocols for V2X Environments

It is widely acknowledged, that one of the key factors to the successful establishment of VSNs and the effective and efficient function of various applications/services, is the design and development of *routing protocols*. However, due to the inherent characteristics of the vehicular environment, traditional routing protocols for MANETs cannot be applied to VANETs. During the past years a number of new routing protocols have been explicitly proposed for VANETs that take under consideration these unique characteristics and aim to route information among vehicular and road-side nodes.

VADD (Vehicle Assisted Data Delivery) [126] is a unicast, delay-tolerant protocol that uses beacon-driven geographic routing to forward packets from source to destination. It adopts the idea of carry-and-forward based on the use of predictable vehicle mobility, in order to achieve low data delivery delay in sparse networks. Having knowledge of the underlying road infrastructure through a static map, packets are forwarded along streets and routing decisions take place at intersections where vehicles select the next forwarding path (series of consecutive streets) to destination with the smallest packet delivery delay. Through a stochastic model that takes into consideration vehicle density on a road, road length and average vehicle velocity, the expected packet delivery delay can be estimated.

GPCR (Greedy Perimeter Coordinator Routing) [82] is an overlay, non-delay tolerant protocol that also uses beacon-driven geographic routing to forward packets from a source to destination. Similar to VADD, packet forwarding is performed along streets in a greedy manner and routing decisions are taken at intersections. GPCR takes advantage of the fact that streets and intersections form a natural planar graph and uses it as a repair strategy when a packet reaches a local optimum. Key to the routing process is the detection of nodes (coordinators) located on an intersection, since GPCR does not rely on an underlying street map. To do so, GPCR employs two approaches: i) *neighbor tables*, where a node x is on an intersection, if it has two neighbors y and z that are within range of each other but do not list each other as neighbors, ii) a *correlation coefficient* that relates a node with its neighbors with respect to their position.

A delay-bounded routing protocol for VANETs was proposed in [109], which aims at satisfying user-defined delay requirements, while at the same time minimizing the overhead imposed to the already limited network capacity. Specifically, the authors propose two new algorithms, D-MinCost & D-Greedy, that enable the delivery of information between vehicles and Road Side Units (RSUs), using the carry-and-forward paradigm. The two algorithms utilize real-time information and past statistics about the underlying vehicular traffic in order to fine-tune the process of switching between multi-hop forwarding and data muling. Given the vehicle and RSU geographic positions, D-Greedy calculates the shortest path between them and allocates a constrained delay-time to each street, composing that path according to path-length and vehicle density. If messages can be delivered within the user-defined delay constraints, then data muling is chosen so that messages are forwarded at the vehicle's speed. Else, multi-hop forwarding strategies are applied to quickly forward the messages. D-MinCost, considers the global urban traffic information and tries to achieve the minimum channel utilization. According to traffic statistics, the cost and delay of each street can be pre-computed.

LOUVRE (Landmark Overlays for Urban Vehicular Routing Environments) [73] is a geography dependant proactive routing protocol. Its design aims to overcome the traditional problems found in geographic protocols such as the concept of perimeter routing, and provide an obstacle-free geographic routing of messages in a VSN by building an overlay network on top of the physical road topology. Particularly, LOUVRE aims at placing overlay landmarks over road intersections where vehicles exist. Consequently, the vehicular traffic density of the roads between any two overlay landmarks is calculated in a distributed manner; each vehicle in these roads makes an approximation of the density of the road by keeping a record of the number of unique neighbors it has encountered on that road. Additionally, the density of other roads is obtained via the broadcast messages of their neighbors. If the approximated density is above a particular threshold (function of the road length and the hardware communication range), an overlay link is established between the two adjacent overlay landmarks. After an initial boot-strap period, an overlay network between landmarks is built. A routing table maintains the best paths from and to any landmark in the topology. Although contrary to the dynamic nature of VSNs, the authors of [73] support the usage of a routing table at the overlay level, by stating that during times of high-traffic volume, the road density stabilizes and thus all vehicles have a very similar view of the road topology. Consequently, the overlay network topology should remain stable and the routing table will not need very frequent updates.

6.3 Connectivity Analysis using Complex Network Science

There is a rich body of work currently on the literature that deals with scientific frameworks for studying the temporal evolution of a number of real-world systems [79]. These graphs span a broad range of domains (autonomous systems, e-mail networks, citations, etc.) and their study leads to significant implications since most of real-world dynamic networks (online social networks, Internet, etc.) have been proved to follow some topological statistical features (i.e. scale-free networks, small-world properties, power-law degree distribution).

In this PhD thesis, we focus on exploring the time-and-space evolving VANET graphs where the mobility of nodes affects network connectivity over space and time in a unique way. As mentioned in Section 1, IVC is a promising field of research, where advances in wireless and mobile ad hoc networks can be applied to real-life problems (traffic jams [19, 94], road accidents [16, 17] etc.) and lead to a great market potential [2, 11]. Nevertheless, connectivity dynamics determine the performance of networking protocols when these are employed in vehicle-based, large-scale communication systems.

The value of the connectivity analysis of ad hoc networks is so fundamental that a competition-experiment was initiated — the MANIAC experiment [111] — to study network connectivity, diameter, node degree distribution, clustering, frequency of topology changes, route length distribution, route asymmetry, frequency of route changes, and packet delivery ratio. The obtained results show a high degree of topology and route changes, even when mobility is low, and a prevalence of asymmetric routes, both of which contradict assumptions commonly made in MANET simulation studies. Previously, authors in [103] used simulations to study the probability densities of link lifetime and route lifetime for some mobility models. According to this study, the path duration seems to be a good metric in order to predict the general trends in the performance of vehicular routing protocols. Another sound observation is that they showed the relationship between the path duration and other critical parameters such as the transmission range and the average relative speed of the mobile nodes, and the average number of hops in the path. Also, well-known concepts from social network analysis have been used as primitives to design advanced protocols for routing and caching in DTNs and ad hoc sensor networks. In [41], the betweenness centrality index and its combination with a similarity metric (SimBet) have been used to select forwarding nodes to support information routing in DTNs. Results showed that data dissemination is improved if the messages are delivered through nodes which have high SimBet utility values. The betweenness centrality has also been used in [44] to design a cooperative caching protocol for wireless multimedia sensor networks. This protocol selects the mediator nodes that coordinate the caching decisions based on their "significant" position in the network. Likewise, the MaxProp protocol [34] transfers messages based on the mobility of intermediate nodes. In a related study [55], the authors combine short-range communication and cellular communication to facilitate query processing in VANETs. Yoneki studied the impact of connective information (clustering, network transitivity, and strong community structure) on epidemic routing [125].

In the context of MANETs, although research works have been published that deal with the aspects of network connectivity [104, 120], only a few consider network graph analysis. Specifically, in [35], the authors study the temporal evolution of the diameter of opportunistic mobile networks, which follow the random graph model. Results have shown that the diameter increases slowly with the network size. Harri et al. in [57] introduced the concept of kinetic graphs to capture the dynamics of mobile graph structures so as to efficiently support network-wide operations, e.g., broadcasting. A kinetic graph comprises a generalization of the static network graph able to model the trajectories of the mobile nodes and supports the notion of the "probabilistic existence" of graph edges.

In the context of vehicular networking, [39] presents a preliminary characterization of the connectivity of a VANET operating in an urban environment. The authors transform the vehicular network into a transitive closure graph. Then, the temporal evolution of the average node degree is examined. Nonetheless, this work does not perform a deep analysis of the networking shape of vehicular mobility and is limited simply to the average node degree for a small time interval. In [108], the authors set up a real-world experiment consisting of 10 vehicles making loops in a 5-mile segment of a freeway. They focus on the connectivity issues without investigating the topological properties of the VANET graph. Fiore et. al [52] study the node degree distribution, link duration, clustering[1] coefficient and number of clusters for VANET graphs under various vehicular mobility models. The objective of [52] focuses on studying the topological properties of different mobility models and explaining why different models lead to dissimilar network protocol performance. The authors

---

[1] The terms "cluster" and "component" are used interchangeably in this manuscript.

of [66] provide an analysis of the connectivity of vehicular networks by leveraging on well-known results of percolation theory. Using a simulation model, they study the influence of vehicle density, the proportion of equipped vehicles, transmission range, traffic lights and roadside units. Similarly, [90] studies the distributions of node degree and link duration in VANETs using a realistic urban traffic simulator. In a recent study [59], authors study the vehicular network in order to develop a stochastic traffic model for VANETs. This model captures spatial and temporal characteristics of a vehicular network, vehicle movement, link condition, and node connectivity.

Viriyasitavat et. al [116], present a comprehensive analytical framework, along with a simulation framework, for network connectivity of urban VANETs, using some key system parameters such as link duration, connection duration, and re-healing time. As a platform for their study, the authors use data from a cellular automata-based traffic mobility model based on a Manhattan-grid like road topology. The analytical framework leads to closed-form expressions which capture the impact of four critical parameters (network density, transmission range, traffic light mechanisms, and size of a road block) on network connectivity. Monteiro et. al [92] also make use of the above CA traffic model in order to analyze simple network features such as node degree distribution, average shortest path length and clustering co-efficient. By extending a known VANET-dedicated protocol, the authors provide preliminary results that indicate the benefits (network overhead reduction) when using topological information during information broadcast.

Despite the thorough evaluation, the above research works base their observations in the study and analysis of a small-scale VANET deployed on a rectangular-grid road topology. Given that the majority of real-world urban environments do not pertain to the simplistic layout and characteristics of such road topologies, the findings concerning network connectivity properties could be biased [86]. To the best of our knowledge, our preliminary work in [97] was the first effort to study the structural evolution of a large-scale VANET topology embedded in real-world road network. Using the most realistic mobility traces publicly available at the time, we analyzed the VANET communication graph among thousands of vehicles driven in Zurich, Switzerland. Despite the valuable findings of that study, the employed synthetic traces were characterized by highly variable vehicle density and a rather non-uniform traffic distribution, which hindered the precise capture of the topological features. The same Zurich traces are more recently employed in [45] in a quest to find social properties in VANETs. The authors examine macroscopic (distance, diameter, density and edge persistence) and microscopic (degree, cluster coefficient and closeness centrality) metrics to prove or disprove the existence of social behavior that emerge with opportunistic vehicle interactions. Similarly to our findings in [97], they characterize VANETs as scale free, since they identify (during peak-traffic hours) the presence of a degree distribution that follows a Power Law.

However, the authors of [93] advocate that the Zurich mobility dataset characteristics (limited geographic coverage and time-span of Zurich traffic as well as inclusion of major arterial roads only) contribute to a vehicular network which is biased towards high IVC connectivity. To this end, [93] provides a more complete analysis of the instantaneous VANET communication graph in Cologne, Germany. The authors employed the TAPAS-Cologne [24] realistic mobility dataset that describes the mobility of several thousand vehicles moving in the urban and sub-urban regions of Cologne, Germany. The results unveiled that the VANET is composed of a large number of disconnected components that hinder delay-sensitive, multi-hop communication among vehicular peers. Furthermore, it revealed that large components are unreliable in terms of their internal dynamics (frequent re-wiring of multi-hop paths), ascribed to their significant spatio-temporal variability. The VANET also lacks navigability since it does not exhibit the properties of scale-free networks, where high-degree nodes are always present and serve as connectivity mediators among other low-degree nodes.

Although [93] provided new and interesting findings regarding the VANET topology, these were summarized over a 24-hour time-frame analysis, which included both peak and off-peak traffic periods. However, all the fascinating phenomena appear during the morning and afternoon peak-hours, at which time, the prevailing traffic conditions shape the vehicular network in a way that some kind of inter-vehicular communication can effectively take place. In this work, we focus on the VANET communication graph realized during the morning peak-period and we commit to a *fine-grained* analysis of the respective topology features and dynamics by employing additional measurements from complex network science. Furthermore, the VANET can be "seen" from the viewpoint of real-world networks that are embedded in low-dimension geometric structures (e.g Power Grid on surface of the earth). Consequently, besides its obvious temporal evolution, it also exhibits a spatial dimension and should thus be studied accordingly. As a consequence, in contrast to the above research works, we aim to identify where in the underlying road network particular network phenomena occur and which are the factors that drive them.

An interesting discussion can be made in regards to the specifics of the time-dimension. All the above works, including ours, model and observe the VANET as *contact sequences*, that is segment the graph into adjacent time windows where contacts are aggregated into edges, and then study the time evolution of the network structure in these windows. However, such studies can be biased towards the connectivity of the network, since the aforementioned approaches do not cover all aspects of the temporal structure of contact patterns. For instance, the edges between nodes of networks might not be transitive in the course of time. This might cause different characteristics of the network such as redundant paths, diameter, network connectivity to be lost and not captured. Therefore, time ordering of the network and when the edges are active is important. Another approach is to consider dynamic and time evolving networks such as the VANET as a *temporal complex system* and study them using an *interval graph* representation [60].

In turn, interval graphs enable us to study different time-respecting measures. Time-respecting paths indicate the pathways information diffusion processes might consider inside an evolving network. In temporal networks, pathways must be defined as sequences of edge activations with non-decreasing times that connect a number of nodes [67]. In contrast to paths in contact sequences, two nodes might not ever become connected, due to the lack of one or more sequential edge activations. Temporal paths are not transitive, whereas paths from contact sequences might 'look' as transitive, depending on the observation window granularity. The reachability ratio [61], is another interesting measure that quantifies the average fraction of nodes inside the set of influence of all other nodes in the network. An influence set of node i, defines the set of nodes that can be reached by any time-respecting path from node i. The reachability ratio gives insights on information diffusion processes in the network, particularly who is able to received an information item T_i at a particular time instance or later. The reachability ratio, can be utilized in tandem with the measure of information latency [72, 89] for characterizing the 'velocity' of the temporal network, and hence evaluate how quickly nodes can on average transmit and information item to each other.

6.4 Traffic Information Systems

Traffic Information Systems were among the first applications proposed that capitalize the potential of V2X communication. In SOTIS [119] and TrafficView [94] vehicles periodically broadcast their position and speed information in order to enable other vehicles to estimate the existence of traffic congestion. While TrafficView focuses on monitoring the congestion of the road directly ahead, SOTIS extends the idea to both sides of the road. StreetSmart [46] pertains to the above concept but uses clustering and epidemic communication to disseminate traffic information. Simulative and field evaluations have shown that these proactive traffic information dissemination techniques work on small and sparse VANETs. However, this is unlikely to be the case in geographically larger and more dense VANETs, since the amount of traffic data that vehicles will collect and consequently broadcast will increase quadratically [83].

VITP [42] specifies the syntax and semantics of messages exchanged between the software components (called VITP peers) of this infrastructure. In dense networks, proactive protocols like VITP can avoid saturation caused by flooding and diffusion since information is queried on demand rather than being pushed periodically during discrete time intervals.

Gao et al [53] propose an adaptive query evaluation plan based on the structure of the underlying road topology. Specifically, traffic queries can be issued by a source vehicle towards individual target roads. The evaluation plan is a sub-tree of the road topology with its root being the location of the source vehicle and leaf-nodes the query target road. Based on query construction rules known a priori by all vehicles, each vehicle residing on streets between the root and leaf-nodes, can autonomously decide whether it will participate in the query evaluation process and under which role. In addition, control messages are introduced to provide updates to vehicles on changes in the location of the query source.

However, the method proposed in [53] caters only for querying the traffic flow of one *single* target road. Consequently, if we are to apply this method in order to query the prevailing traffic conditions on a number of roads in a composite path towards the destination, the respective number of packets/queries need to be generated and transmitted. In turn, this imposes a significant overhead in the VANET, which subsequently leads to information loss due to packet collisions in the wireless channel [105]. In addition, the frequent transmission of control packets provides an additional overhead to the already limited network capacity. Given that in a real urban environment a number of road-paths need to be monitored, it is evident that such an approach does not scale.

Connectivity requirements for V2V-based TIS are determined using an analytical framework proposed in [98]. This particular framework enables network designers to determine the minimum penetration ratio of vehicles equipped with IEEE 802.11p hardware necessary to form a connected VANET. Furthermore, it allows to identify the critical transmission range required for this connected network in order to facilitate traffic information distribution in an extended region or over multiple hops. However, the system model employed in this work assumes that vehicles are randomly and uniformly distributed along a road segment, an unrealistic assumption in large-scale environments. In addition, vehicles are assumed to maintain a constant speed while they are driven on a road segment, a factor which causes network connectivity to be biased towards reliability.

To overcome the VANET challenges while disseminating traffic information, Adaptive Traffic Beaconing (ATB) [110] proposes a beacon-based message protocol for message exchange that maintains a congestion-free wireless channel. As its name suggests ATB intelligently adapts the beacon interval period by complimentary monitoring the channel quality and the utility of each broadcasted message. Overall channel quality is assessed by means of three metrics: number of collisions in the channel, signal-to-noise ratio, neighbourhood size. Flexible weights are given to each metric to adjust the protocols reactivity. Message utility is also a compound metric derived by the geographical distance of a vehicle to a traffic incident and also the message age.

However, estimating the prevailing road traffic conditions however is not a trivial task. MobSampling [54] proposes a distributed V2V-based method for vehicular density estimation in an target geographic region. The estimation procedure involves two endpoints: (i) a source vehicle, which queries the traffic density in a remote target area and (ii) a destination vehicle in the target area (a geographic coordinate plus a buffer region), which estimates the density. The source vehicle generates a "sampling kit" in the form of a broadcasted data message. The kit is geographically routed towards the target area and the first vehicle residing in the target area (called the Sampler) that receives the kit, is responsible to provide a traffic density estimate for a time period. The Sampler, starts polling its neighbors for the period of time specified in the kit header so as to obtain a local estimation of vehicle density in its lane. Once the estimation is completed, it can be returned back to the source node via geocasting, or distributed to vehicles in the VANET.

VCAST [71] maintains the use of broadcast messages for traffic information. However, in contrast to MobSampling, it targets multi-hop traffic estimation. Nevertheless, forwarding traffic information over many hops is not scalable since it can impose additional communication overhead. As the vehicle density and area under study increase, then the allowable broadcast rate will decrease to avoid saturation of network resources. To overcome this problem, VCAST exploits the notion of a distance-sensitivity in its traffic propagation information scheme: information about each vehicle is propagated at a rate that decreases linearly with the distance from the vehicle. In addition to actual vehicle location, VCAST can propagate aggregate traffic information (i.e average vehicle speed, density of individual traffic regions etc).

The efficient use of available bandwidth and resource saturation, is also the priority of FairAD [107]. The authors propose the enhancement of the ATB [110] protocol discussed above, with the concept of Nash Bargaining from game theory. This extension which works in tandem with broadcast rate control, allows the dissemination protocol to distribute traffic data fairly over vehicles while adaptively controlling network load.

6.5 Dynamic Information Dissemination

The design and development of data dissemination mechanisms for VANETs, which aim at increasing the ratio of solved traffic information queries with the minimum network overhead, while maintaining acceptable levels of information quality, has been the target of research investigations in the literature [50, 53].

Perhaps the first effort to study caching in vehicular networking infrastructure was presented in [51] with the Infoshare application. Infoshare provides a cooperative caching approach in VANETs. When vehicles request a piece of information, the information is cached to the querying vehicle for a certain time interval. However, this application is limited by the used transport protocol since the proposed scheme does not support any cache-based protocol for inter-vehicular communication. Furthermore, the authors examine the impact of caching in VANETs using a simple straight-road mobility scenario. From a technical point of view, in urban environments we have more complex mobility scenarios (e.g., alternate routes, traffic jams, junctions etc.) than what we have in straight or highway roads. Therefore, it is crucial to study the effects of caching in such environments; cache-based location-aware services in VANETs may reduce the automobile traffic. More recently, a trajectory-based data forwarding scheme, called TBD, for light-traffic road networks has been proposed [65]. This scheme exploits both private trajectory information and traffic statistics.

Fiore et. al. in [50] presented Hamlet, a fully-distributed caching scheme in which vehicles decide independently of each other whether to cache and for how long a piece of information. Decisions are made based on individual node's observation of the information present within its radio range. The objective of this approach is to increase content availability in the proximity of nodes and minimize network overhead. Although Hamlet demonstrates an increase in the query success rate and a decrease in the network overhead, the results are based on a confined static information data-set within a simple urban mobility scenario. However, it is important to study the effects of caching in urban environments with diverse topographical layouts using a larger information data-set, since the performance of caching in VANETs is mainly affected by the mobility scenarios. Nevertheless, the authors do not consider a cache-based, proactive, communication protocol for the dissemination of vehicular information in their experiments, nor do they conduct experiments in dense mobility scenarios where information is generated by both mobile and stationary nodes and under the presence of unscheduled events like accidents.

In [87] the VTube autonomous and cost-effective infrastructure is proposed to facilitate low-cost and QoS guaranteed information publish/subscribe in an urban environment. It relies on geographically distributed fixed RSUs installed in various facilities (i.e restaurants, gas stations, stores) to cache and publish content for vehicular users. Vehicles are used to spread content among geographically separate RSUs and also serve as mobile caches. The goal of VTube is to enable collaborative caching among RSUs and provide guaranteed download performance to users in terms of download delay constraints. Files are fragmented into smaller pieces of equal length, and vehicles could finish downloading a file when they collect all pieces. Files are stored in RSUs stochastically to improve the download delay. The file replication in RSUs is a function of the respective popularity and availability of files in the VSN.

Chapter 7

Experimentation and Evaluation TestBed

This section presents a multi-level experimentation testbed, which was established in order to evaluate the different research approaches and methodologies proposed for solving the problems presented earlier in Section 3.

These include the different mobility datasets utilized throughout this thesis and the different software packages developed. For the latter, standard object oriented programming (OOP) methodologies were employed for the majority of the developed software, mainly in Java and C++ programming languages.

Various scripting language including Perl[1], Bash[2] and Python[3] were utilized to handle several tasks such as data processing and cleanup, code distribution and execution scheduling, result collection and visualization. In addition, all source files include the minimum language specific documentation that describes the functionality of the code or any respective application programming interfaces (API).

---

[1] http://www.perl.org

[2] https://www.gnu.org/s/bash/bash.html

[3] https://www.python.org/

For consistency and backup reasons, all software packages that comprise the evaluation testbed were maintained in the high-performance computing and storage infrastructure provided by the Laboratory for Internet Computing (LINC). [4] The available versioning service (SVN) was employed for storing, keeping track of chronological changes in each of the individual source files. Code was committed on a nightly basis to the SVN server.

All software packages were developed following the open-source model, with the aim of making them available to the vehicular transportation community. To this end all software packages presented in the section are publicly available through GitHub in [5].

7.1 Datasets

The structure and evolution of the vehicular network, as well as its capacity to sustain any level of V2X information diffusion, is primarily dependent on the mobility dynamics of vehicles, as well as the location of fixed road-side units (RSU). Mobility dynamics are a function of a number of interrelated parameters including the driver behavior, the structure and constraints of the underlying road topology, traffic light signalization, diurnal processes and even people's societal habits. Datasets with mobility traces are key to the study of any VANET-related since they provide a spatio-temporal profile of vehicles' movement either in an urban, sub-urban or highway environment. A quest for this work was to identify and utilize such mobility traces that not only provide accurate but also high fidelity characterization of vehicles' mobility in real-world large-scale urban environments. Except of real traces, we also study traces derived from mobility models which are proven to be realistic. These datasets were then utilized extensively during the VANET Evolutionary Dynamics, Urban Traffic Sensing and Acquisition and Traffic Information Diffusion studies. The

---

[4] http://linc.ucy.ac.cy/index.php/infrastructure

Hermes Moving Object Database (HermesMOD) [9] was utilized in order to store all the datasets and also being able to quickly index and query the different moving objects (vehicles). HermesMOD allowed us to get knowledge about the vehicle movement behavior and generate various mobility statistics. Table 7.1 provides an overview of all the mobility dataset cataloque.

	Area Size	Vehicle Density (veh/lane/Km)	Average Speed (m/s)	Total Vehicles	Recorded Time
Real					
Nicosia	$35Km^2$	-	15.8	15	4 Months
Shanghai	$4Km^2$	-	4.72	700	2 Hrs
CabSpotting	$9Km^2$	-	13.8	472	3 Hrs
Synthetic					
Nicosia	$8Km^2$	7	11	970	15 Mins
Los Angeles	$4Km^2$	10	11.2	7000	2 Hrs
Cologne	$33x35Km$	8	13.8	75579	2 Hrs

Table 7.1: Mobility Dataset Overview

7.1.1 Real Datasets

Nicosia Dataset: Real mobility traces from the Nicosia area were obtained by logging the travel patterns of 15 individuals. Six HTC smartphones were acquired and given to each person (in random order) so as to collect real mobility traces for a period of 4 months. Each phone logged the daily mobility of each individual and then uploads it to a private trajectory repository through WiFi or 3G cellular network. For logging users mobility traces, an open-source, publicly available application called GPSLogger [5] was utilized.

The GPSLogger application was customized in order to enable WiFi access preference over 3G for uploading mobility traces in order to minimize financial expenses. Due to privacy concerns, access to these mobility traces is not publicly available, so as to guarantee that the sensitive information related to each individual (identity,

---

[5] http://code.mendhak.com/gpslogger/

Started
Jul 26, 2012 9:23:15 PM (Using GPS satellites)
Latitude: 37.422005
Longitude: -122.084095
Altitude: 0.0 meters
Speed: n/a
Direction: n/a
Satellites:
Accuracy: n/a

Figure 7.1: Real Mobility Traces from Nicosia using GPSLogger

home and work locations etc.) will remain private. We will examine ways in order to provide such traces to the research community in the future.

Shanghai Taxi Dataset [18]: Real mobility traces were obtained through the Smart City Research Group. SCRG maintains a dataset of real GPS traces collected from taxis traveling throughout Shanghai, China, in a 24-hour time period. We developed a tool in order to parse the taxi traces and converted them from GPS cylindrical coordinates to Cartesian plane coordinates for better manipulation. A $2Km \times 2Km$ rectangular region around Shanghai city-center was isolated, and the mobility traces of all the taxis that run within this region for a 2-hour time period, were utilized. This clipping process resulted in obtaining the mobility traces of 704 distinct vehicles for our study. According to the authors' knowledge, the SCRG data set is the largest, publicly available dataset for vehicular traces.

CabSpotting Dataset [22]: Real mobility traces were also obtained through the CabSpotting.org project. CabSpotting collects real GPS traces from San Francisco's taxi cabs as these travel throughout the Bay Area. The conversion tool developed in [70] was used to transform the CabSpotting traces from GPS cylindrical coordinates to Cartesian plane coordinates for better manipulation. A $3Km \times 3Km$ rectangular region surrounding the San Fransicso Bay Area center was isolated, and the mobility traces of all the taxi cabs that run in this region for a 3 hour time period, were extracted. This clipping process resulted in obtaining the mobility traces of 432 distinct vehicles (taxi cabs) for our study.

Road-Side Unit Dataset: The increasing deployment of inexpensive WiFi hotspots in and around urban environments could be beneficial for VANET's. With effective communication with moving vehicles being established and maintained up to 300 meters , such stationary Road-Side Units (RSU's) can improve network-wide connectivity by acting as bridges between two or more fragmented regions during the initial VANET deployment and low market penetration phases. Furthermore, even when high market penetration is achieved, such RSU's can act as information service providers or Internet gateways. We study extensively how the presence of RSUs influences the structure and evolution of the VANET communication graph. To achieve this, we extend the available datasets with stationary RSUs. The Wigle.net [29] online catalog was utilized in order to extract the position of Wi-Fi hot-spots in the area of study, which potentially could be utilized as RSUs. Wigle maintains a publicly available and constantly user-updated catalog of wireless networks throughout the US, with each network attributed with identification (SSID), location (latitude/-longitude), and security (open/private) information. We developed a tool to extract all open Wi-Fi Access Points (AP) in the area of study and placed then on the map by converting the GPS cylindrical coordinates to Cartesian plane coordinates.

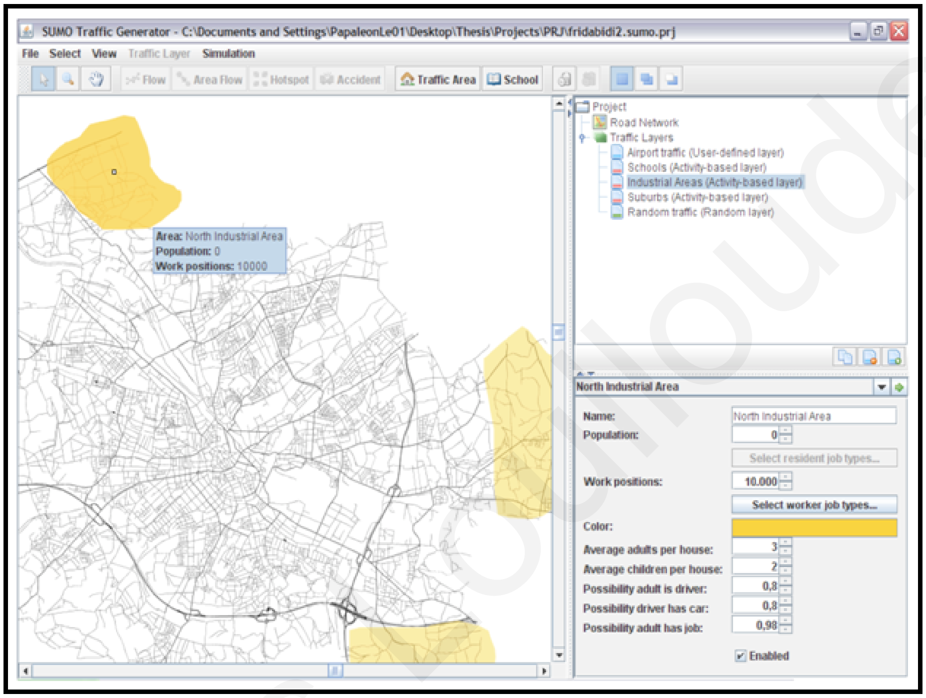
7.1.2 Synthetic (Realistic) Datasets

Due to the size constraints, and scaricity of real traces, we enhance our dataset collection evaluation by studying syntetic mobility traces in real urban road topologies. For this purpose, we employed well-known traffic simulators or obtain publicly available mobility traces, aiming always for the *highest level of realism*.

Los Angeles Dataset: Mobility traces were generated using the VanetMobiSim [58] vehicular mobility generator. By employing known traffic generation models, VanetMobiSim outputs detailed mobility traces over real-world, accurate, city maps available in the Topological Integrated Geographic Encoding and Referencing (TIGER) database [27] from U.S. Census Bureau. The realism of the mobility models utilized in VanetMobiSim has been validated extensively in [58] using benchmark tests from vehicular traffic flow theory which are highly accepted by the transportation community.

In particular, we generate vehicular traffic within a $2Km \times 2Km$ area bounding the city center of Los Angeles, CA. In the generated scenario all vehicles were set to follow the Intelligent Driver Model with Lane Changes (IDM-LC). More specifically, IDM-LC falls into the car-following mobility models category since individual vehicle behaviour depends on the behaviour of the preceding vehicle. It extends the basic IDM model, in the sense that it adds intersection handling capabilities to the behaviour of vehicles and the possibility that vehicles change lanes and overtake other preceding vehicles. By examining the respective TIGER map, VanetMobiSim was set to synchronize the operation of traffic signals of 110 different intersections.

To achieve an even higher level of realism in the vehicular traffic generated, we opted to define the average vehicle density (and hence the total number of vehicles) for the area under study, by examining the 2010 Annual Average Daily Traffic (AADT) statistics, which are publicly available through California's Department of Transportation Traffic Data Branch [26]. AADT measures the total volume of vehicle traffic on a given road for a year divided by 365 days. In general, AADT is obtained through traffic counting using electronic equipment installed on the roadside. Our study on a large number of roads in the area of interest, indicated an average density of *10 veh/lane/Km* during normal driving conditions (not peak hours). With approximately 700 Km of road length (one-way and number of lanes for each individual roads included), VanetMobiSim was set to generate vehicular traffic for a 2 hour period, constantly maintaining 7000 vehicles in the area of interest.



Area: North Industrial Area
Population: 0
Work positions: 10000

Name: North Industrial Area
Population: 0
Work positions: 10.000
Average adults per house: 3
Average children per house: 2
Possibility adult is driver: 0,8
Possibility driver has car: 0,8
Possibility adult has job: 0,98
☑ Enabled

Figure 7.2: TrafficModeller - Generating Synthetic traces for Nicosia

Nicosia Dataset: Synthetic mobility traces for Nicosia were generated with the help of TrafficModeller [99]. The map was extracted from real-world, accurate city topologies obtained from OpenStreetMap [14]. Furthermore, using TrafficModeller again, a number of "hot-spots" within this region were defined (see Figure 7.2), in order to simulate the traffic conditions that arise when people drive from/to their workplaces, shopping malls, amusement centers etc. Vehicles in the simulation were set to have an acceleration of 4.5 m/s, deceleration of 2.6 m/s with their top speed bounded by the road speed limit. Driver imperfection factor, that is the ability of the driver to adapt to a desired speed, was set to 0.5 (where the value 1 indicates a perfect driver). Total simulation time was set to 1000 seconds. These high-level abstractions were translated as low-level input specification and passed to SUMO. Vehicular traffic traces for 970 distinct vehicles that move inside the region of interest were generated by SUMO. Vehicles have a mean drive time of approximately 230 seconds. The average vehicle speed is between 8 and 10 m/s. Vehicle average speeds have a large variance but follow a uniform distribution. Mobility traces obtained from TrafficModeller and SUMO, were transformed to acceptable ns-3 input trace files using TraceExporter [25].

TAPAS-Cologne Dataset [24]: provided by the Institute of Transportation Systems at the German Aerospace Center (ITS-DLR). Based on real census information concerning the daily commute patterns of people in the city of Cologne (Germany), the TAPAS-Cologne project provides an extensive dataset of mobility traces that describe, with very high realism, the vehicular traffic within and around the city of Cologne for a period of 24 hours.

Figure 7.3: City of Cologne Road Network Topology

In this work, we consider a corrected and enhanced version of the initial TAPAS-Cologne dataset, following the results of [115]. This version provides the road traffic for the early morning rush-hour in Cologne from 6:00 to 8:00 a.m. The level of realism of the particular dataset is acknowledged by the research community and has recently been employed in [93] to assess the level of vehicular connectivity in large-scale urban environments. Figure 7.3 provides a screenshot extracted from the OpenStreetmap (OSM) [14] online mapping service, depicting the road network connecting the city of Cologne and the surrounding suburbs, included in the TAPAS-Cologne dataset. The dataset covers approximately an area of $33 \times 35Km$ with 42148 intersections and 134645 roads. We employ information from the OSM database, to derive the following taxonomy concerning the core features of the road-network in the City of Cologne, namely roads and intersections. The aim behind this process is to identify whether particular phenomena in the VANET emerge in different features of the road network and which attributes of such features affect their spatio-temporal evolution.

- *Roads*

  1. *Urban Roads*: Low capacity roadways, which provide access to residential areas with houses or various amenities on either side. Urban roads can be one-way or two-way and in their majority have only 1 lane per travel direction.The maximum speed allowed on such roads is 30 Km/h (8.3 m/s). Urban roads appear in white color in Figure 7.3.
  2. *Collector Roads*: Low-to-moderate capacity roadways designed to collect traffic from urban roads and carry it to arterial roads or vice-versa. Traffic flow is mostly managed via the utilization of signalized intersections, round-abouts and stop-signs. Collector roads have 2 lanes per travel direction. Since these roadways are located inside build-up areas, the maximum

speed limit varies between 30 to 50 Km/h (8.3 to 13.9 m/s), depending on the area. Collector roads appear in yellow color in Figure 7.3.

3. *Arterial Roads*: High-capacity, bi-directional urban roadways that are used explicitly to deliver traffic between collector roads and freeways and also to carry long-distance traffic flows between different activity areas. They are considered as the back-bone of the road network and traditionally are arranged in concentric circles. Arterial roads have 2 lanes in each drive direction and usually are separated into two classes: (i) Low-speed, where the the maximum speed limit is 50 Km/h (13.9 m/s) and (ii) high-speed, where the maximum speed limit is 80 Km/h (22.2 m/s). In Figure 7.3, arterial roads appear in red color.

4. *Freeways*: Multi-lane, bi-directional, high volume roadways which are dedicated to high speed vehicular traffic. In such roadways, traffic flow is not obstructed by any stop signs and traffic signals. Separation and connection with other roadways of lower grade (i.e arterial and collector roads) is carried out via overpasses across the highway span. Entrance or exit to freeway are provided at overpasses by ramps that allow speed transitions between the freeway and lower grade roadways. Freeways have 2 lanes in each drive direction and the maximum speed limit is 120 Km/h (33.3 m/s). Freeways appear in blue color in Figure 7.3.

- *Intersections:* Intersections are classified depending on the number of road segments that are adjacent to the intersection itself and also the traffic control method. Since the aforementioned dataset lacks controlled intersections using traffic lights [6], we classify them using the adjacent road segments. Using an automated parsing of the Cologne road network, we have identified 3-way ("T"

---

[6] All intersections are stop/yield-sign controlled

or “Y” shape), 4-way (crossroads), 5-way and 6-way intersections. Additionally, we identify intersections in the form of roundabouts with 3 or more exists.

7.2 VEGA: An Extensible Framework for Large-Scale V2X Graph Generation and Analysis

As discussed previously, exploring the behavioral dynamics of inter-vehicle communication in urban environments from the perspective of Complex Network science, requires modelling the VANET as a graph, wherein vehicles correspond to the set of vertices and wireless ad hoc links correspond to the set of edges. Taking under consideration that VANETs are an evolving network, both in space and time, the modelling process can be extended by representing the network as time sequence of graphs (graphlets) wherein each node is embedded in a low-dimension coordinate space. Depending on the time resolution and the overall time window of study, the modelling process can result to hundreds or even thousands of graphlets. Understandably, in order to derive conclusive remarks about the spatio-temporal properties of VANETs, one has to establish and follow a well-defined graph generation and analysis workflow. Figure 7.4 provides a high-level view of such a workflow, decomposed into a sequence of 4 discrete activities. Each activity is responsible for the completion of a particular action and requires the presence of a number of heterogeneous software packages that in the end must be made to interface with each other seamlessly.

Starting from the left of Figure 7.4, mobility datasets (real or synthetic) need to undergo a pre-processing step and get converted to suitable input for network simulation packages such as ns-3 or Veins [23, 28]. Although *Activity 1* seems trivial at first sight, it requires the development of custom-made programs that understand and convert data between a variety of input/output formats, are able to cope with the noise information usually present in raw mobility datasets, execute coordinate

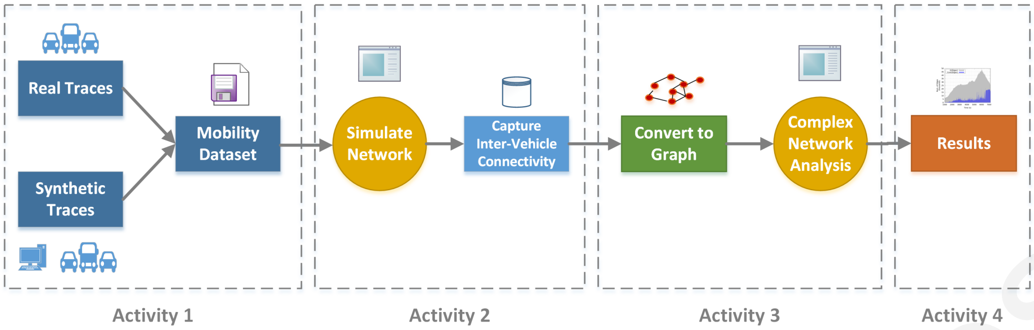


Figure 7.4: VANET connectivity graph generation and analysis workflow

transformations if necessary, and perform trajectory extrapolations in case of missing traces points.

Once mobility datasets are prepared accordingly, *Activity 2* uptakes the task of simulating all aspects of communication in a VANET that can potentially be established over an urban, suburban or even highway setting. By considering the moving patterns of each vehicle - as these are represented by the respective traces in the associated mobility dataset - network simulation packages capture the specifics (source-destination pair, time, location, data packets) of any information exchange over the wireless ad hoc channel. This stage is both time consuming and computationally expensive, given that network simulators tend to be thorough and quite realistic when simulating the internals of each layer in the communication stack (physical, link, network, transport, application). In the case where the experimentation scenario dictates the study of large-scale and dense geographic areas with thousands of active vehicles (i.e highly populated metropolitan areas), simulations can take days to complete and eventually result in massive amounts of communication trace data. Furthermore, the majority of available simulator packages have minimal inherent support for VANETs, therefore considerable software development effort is required to enhance them (routing protocol development, application workload generators,etc) [86]. On top, extensive familiarization with the simulation package itself is necessary in order to properly configure and correctly execute a VANET simulation scenario.

Activity 3 is a two-step process, key to the whole workflow since it is responsible for generating and analysing the actual VANET connectivity graphs. The first step involves, parsing the simulation traces, identifying and extracting wireless communication synapses, and subsequently converting them in an acceptable graph notation that can be understood by a network analysis toolkit such as JUNG, SNAP, igraph, Gephi, Pajek [3, 10, 12, 15, 20], etc.

Once the VANET is converted in an appropriate complex network structure, the second and most important step is the actual analysis of the generated graphs so as to examine different graph theory measures (i.e network diameter, shortest paths, node importance, component/community properties, etc). The complexity of the process depends on the number of graphlets to be analyzed and subsequently the set of measures to be calculated upon them. The more measures needed to be investigated, the more data transformations and specific algorithm invocations are necessary. The complexity increases even further since not all graph exploration platforms feature the same analytic toolsets, therefore a combination of two or more platforms might be essential to complete a particular study. Additionally, given that VANET communication graphlets are realized from the movement and opportunistic interactions of vehicles while being driven on the underlying road network, they should be "seen" from the viewpoint of networks that are well-embedded in low-dimension geometric structures (e.g Power Grid on surface of the earth). Therefore, the latter introduces the need of applying also a spatial context to network measures in order to be able to identify where and/or why particular phenomena occur on the road network (i.e, which parts of the city do geodesics usually traverse?).

Similarly to Activity 1, this task assumes the development and existence of custom-made tools that are responsible to: a) correctly convert simulation output to an appropriate graph format that describes complex networks structures and their associated data, b) sort and orchestrate the analysis execution of all available graphlets, and c) retrieve, organize and spatialize the resulting network measures to a coordinate space of preference (i.e Euclidean or World Geodetic System).

Finally, *ctivity 4* is responsible for aggregating, performing statistical analysis and rendering the complex analysis results in meaningfull graphical representation(s) that can ultimately convey interesting insights about the spatio-temporal dynamics of the particular VANET scenario under study. Key to this activity is the utilization of data plotting utilities like Gnuplot [7], the selection and refinement of appropriate charts.

From the discussion above, one can easily observe that the workflow entailed until reaching a state where meaningfull and conclusive remarks regarding the evolutionary dynamics of the network can potentially be derived, is both time-consuming and error-prone. *VEGA was designed to solve the above problem by providing an extensible framework based on the Java platform that allows researchers to quickly generate and analyze connectivity graphs for V2X scenarios that span thousands of vertices and edges.* In a nutshell, it automates and streamlines the modelling of time-ordered V2X connectivity graphs on which a host of different complex network analysis measures can be orchestrated for execution through a simple user-friendly configuration. Off the shelf, it supports a large number of mobility trace formats on which to connectivity models can be applied in order to generate the graphlet sequence. Nevertheless, its modular architecture allows researchers to easily plug-in their own custom modelling and analysis components by extending a set of well defined abstractions and core classes. In turn, this provides them with the flexibility of either utilizing simple models that avoid time-consuming computations or utilize fine-grained models - as those encountered in network simulation - effectively produce high fidelity and more accurate results. Furthermore, all the activities in graph generation and analysis workflow (similar to Figure 7.4) are transparent from the user and the only interaction required in the initial configuration and later on the retrieval of the end-results (charts, statistics, etc), Lastly, VEGA incorporates an intelligent thread pooling mechanism that provides efficient utilization of modern multi-core computer architectures, thereby resulting to faster analysis of large-scale V2X scenarios. The following section presents the architecture of VEGA in more detail.

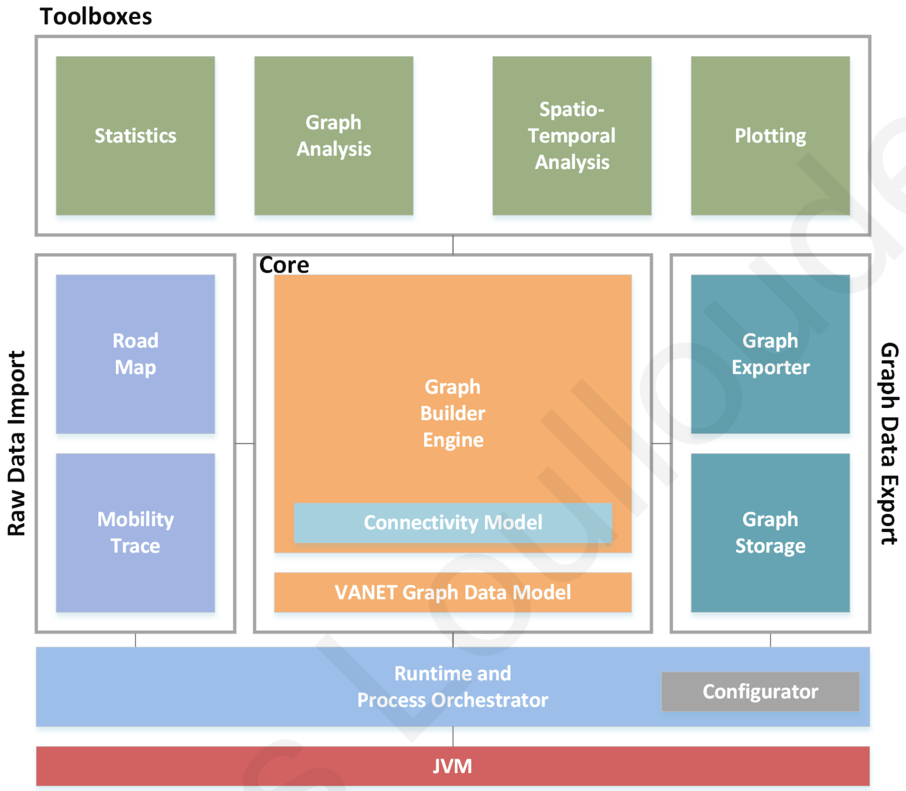


Figure 7.5: High-level view of VEGA framework architecture.

7.2.1 System Architecture

As aforementioned, VEGA adheres to a modular system architecture that is based on the Java Virtual Machine (JVM), therefore it can be executed from the command line interface of any Java-enabled system. It makes use and extends a number of scalable data structures readily available in the Java platform, while taking advantage of native concurrency mechanisms for efficient computing resource utilization. This section presents the different components of the VEGA framework modular architecture, as depicted in Figure 7.5.

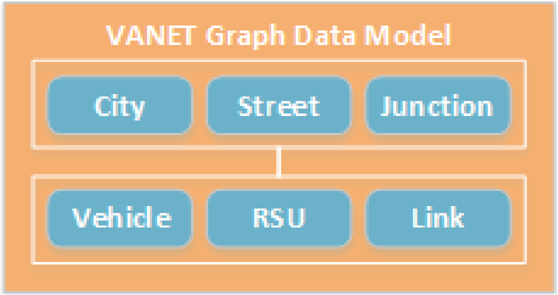


Figure 7.6: VEGA - VANET Graph Data Model Component

VANET Graph Data Model: Provides the necessary abstractions that characterize the presence and movement of vehicles, as well as the establishment of opportunistic V2V and V2I communication in an urban environment. Initially, it includes a sub-layer that encapsulates the base classes for each main actor of a V2X network, particularly *Vehicles, Road-Side Units (RSU)* and opportunistic *Wireless Ad Hoc Links* that can be formed among the former. Base classes include individual attributes - representing the spatio-temporal status of their respective instances - which can be leveraged through specific programmatic interfaces by other VEGA modules in order to construct the all important V2X connectivity graphs. Examples of such attributes are: {*Vehicle*: V_{id}, speed(t), location(t), status(t)}, {*RSU*: U_{id}, location(t), status(t)} and {*Link*: endpoints(V_{id1}, V_{id2}), duration(t), type[V2X | V2I]}.

In addition, the VANET data model includes a second sub-layer for road network related abstractions such *Intersections*, *Streets*, *Light Signalization* and *Traffic Statistics* that enable the characterization of the underlying street-level topology on which vehicles run along or across. As previously, base classes in this sub-layer also include attributes that can be utilized by other VEGA modules for V2X modelling or analyses purposes. Specifically it provides access to the following road network constructs and their attributes: {*Intersection*: I_{id}, position, signalized[T | F], signal_timing}, {*Street*: S_{id}, endpoints(I_{id1}, I_{id2}), street_type, lanes, speed_limit, traffic_statistics}.

Moreover, through a sub-layer connectivity mechanism, the VANET Graph Data Model assures that bi-directional access to the data of each sub-layer is available from the other sub-layer (road topology information can be accessed through V2X class instances, and vice-versa). For instance, the sub-layer connectivity mechanism enables to easily pinpoint the street a particular vehicle runs along, or which vehicles are in the surroundings of an intersection at a specific time instance.

Runtime and Process Orchestator: Responsible to setup the execution environment, manage the VEGA runtime process and also orchestrate the different steps that are necessary in order to model the V2X connectivity graph and generate analyses results. Essentially, the role of this component is to oversee and assert that all the activities in the graph generation and analysis workflow (as in Figure 7.4) are executed in the correct sequence, have and produce the necessary input/output data, respectively. In addition it is responsible to hold the state of different components, log any warnings/errors that are encountered during workflow execution, prompt the user with informative messages about different actions and allocate handles to computing and storage resources.

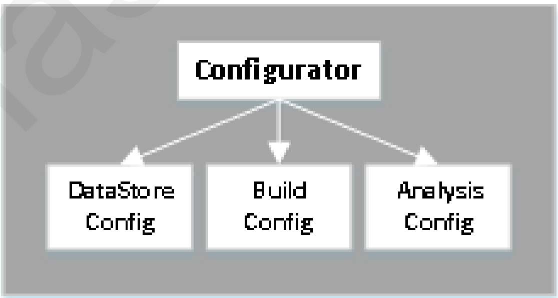


Figure 7.7: VEGA - Configurator sub-component

The task of the *Configurator* sub-component is to setup the execution environment, by consuming small user-provided properties files containing key-value pairs that guide the VEGA Runtime and orchestrator regarding which processes should be invoked and their parameters thereof. As depicted in Figure 7.7, setup of the execution environment is divided in three key operations: (i) *DataStore configuration*, providing information about the location and format of various input and output data, (ii) *Build configuration*, providing information regarding V2X graph build process such as the connectivity model to be utilized, mobility trace format, the overall time window and resolution (i.e snapshot granularity), V2X technology penetration, graph export format, (iii) *Analysis configuration*, providing information regarding the V2X connectivity analysis process such as the complex network metrics to be produced, the measurement granularity, and how results and statistics will be reported to the user (charts or tabular presentation).

```
config.operations = PARSE,BUILD,ANALYSIS

# Root Input/Output Directories
input.root.dir = <userdir>VEGA/TAPAS/input/
output.root.dir = <userdir>VEGA/TAPAS/out/
exec.root.dir = <userdir>VEGA/TAPAS/exec/

log4j.properties = $exec.root.dir /Conf/log4j.properties
datastore.config.file = $exec.root.dir/Conf/DataStore.conf
build.config.file = $exec.root.dir/Conf/Build.conf

#Road Network Information
use.roadnetwork = true
roadnetwork.type = OSM
roadnetwork.file = $input.root.dir/Map/cologne.osm
offset.x = -342498.94
offset.y = -5630725.14
```

Listing 7.1: VEGA Base configuration

```
# Datastore type [DB | FILE]
datastore.type = DB

# DataBase Storage Details
dbase.hostname = localhost
dbase.port = 3306
dbase.name = urban_v2x_overlay
dbase.user = <user>
dbase.pass = <pass>
```

Listing 7.2: VEGA DataStore configuration

```
# Parse a SUMO mobility trace file
trace.format = sumo
trace.sumo.fcd.file = $input.root.dir/SUMO/fcd.xml
trace.sumo.edgestate.file = $input.root.dir/SUMO/agg-edges.run.xml
trace.sumo.netstate.file = $input.root.dir/SUMO/netstate-dump.xml

# Connectivity Model to be used
connectivity.model = LOUVRE

# Transmission Range
tx.los = 250
tx.nlos = 140

# Penetratio Ratio
pratio.low = 0.2
pratio.high = 1.0
pratio.step = 0.2

# Type of graph [static (default) | temporal]
generate.graph = static

# Get the snapshots every 10 second
generate.snapshots = true
snapshot.destination = $output.root.dir/Graphs
timestep = 10

# Snapshot Output Type
snapshot.output.type = DB
```

```
30 # Starting second
31 low.time = 21600
32 # Ending second
33 high.time = 28790

35 # Snapshot Exception Window
36 exception.window = false
37 window.low.time = 26650
38 window.high.time = 26750
39 window.time.step = 1
```

Listing 7.3: VEGA Build configuration

```
 1 # Analysis Time Step
 2 timestep = 10
 3 # Starting second
 4 low.time = 21600
 5 # Ending second
 6 high.time = 25200

 8 # Analyses to performed on the Graph Files
 9 analyses = NETWORK_DIAMETER; DEGREE_DISTR; COMMUNITY_SIZE; BETWEENESS_CENTRALITY;

11 plot = GNUPLOT

13 # Export Graph
14 export.graph = true
15 graph.file.type = PAJEK
16 graph.file.prefix = pajek_
17 graph.file.extension = .net
```

Listing 7.4: VEGA Analysis configuration

Listings 7.1, 7.2, 7.3 and 7.4 above provide a brief example of configurations required by VEGA. The base configuration instantiates the necessary properties for the different types of workflow activities (trace parsing, graph building and analysis) to be performed and location of root input/output directories (lines 1–6). In addition it provides links to the remaining configuration files (lines 8–10), as well as optional information about the road network to be used in the forthcoming build and analysis activities (lines 12–17). Next, the datastore configuration provides the runtime with important information regarding where to find input mobility trace files, either by the local file system or from a database, and in the case of the latter any required schema and access details (lines 5–9). The build configuration provides the type of mobility traces and associated files that can be used to enrich the raw data (lines 2–5). In addition it indicates the connectivity model to be used for building V2X graphs in accordance with technology-related parameters such as transmission range and penetration values (lines 7–17). Importantly the build configuration dictates what graphs will be created - that is static contact sequence versus temporal interval graphs, the snapshot granularity (lines 20–25), start and end time instances (lines 31–33), and exception windows where the granularity might be finer for more in-depth network study (lines 36–39). The analysis configuration indicates the graph analysis start-stop temporal instances and frequency (lines 1–6), the specific analyses to be executed on all graphs and how the results will be plotted (lines 9–11) and whether each graph will exported in a standardized graph notation (lines 14–17) for possible further use in other 3rd party tools. Finally, it is worth noting that all configuration properties are globally available for each VEGA module and its operations.

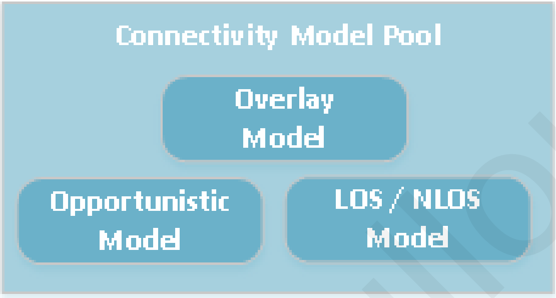


Figure 7.8: VEGA - Connectivity Model

Connectivity Model: This component dictates the fundamentals of how (and if) different nodes in the V2X graph will become connected with each other, effectively guiding the whole topology formation. Each implementation of a specific connectivity model implements a set of well defined interfaces that are realized by specific implementations, which in essence enable the Graph Builder engine to assess whether two actors 'can connect' with each other. Connectivity can depend on a number of simple user-provided parameters such as the nominal transmission range/power of each different type of actor, signal fading rate, global penetration ratio, etc. Currently, as depicted in Figure 7.8, VEGA supports 3 connectivity models: (i) Opportunistic model, (ii) Line-of-Sight (LOS) / Non-Line-of-Sight (NLOS) model, and (iii) Overlay model. The opportunistic model examines the distance between two neighboring nodes and connects them given that their geographic separation is less than a predefined value. The LOS/NLOS model, considers the structure of the underlying topology (provided through the road map) to determine whether two nodes have line of sight or not. Two nodes become connected, if their respective LOS/NLOS value are equal or below a predefined value which is provided by the configuration file. The overlay connectivity model, instead of connecting individual vehicles, connects neighboring road segments given vehicle density on the road. The graph in this model assigns a vertex to each road in the street topology, if the vehicle density on that road is above a particular threshold, again provided through the configuaration file. An edge is drawn only among existing neighboring vertices.

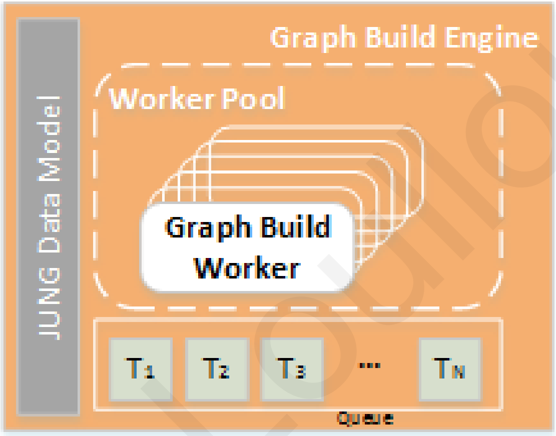


Figure 7.9: VEGA - Graph Build Engine

Graph Build Engine: The build engine is responsible to build all the V2X graphs based on the imported mobility traces and the communication model of preference. The internal representation of the connectivity graph is created, manipulated and stored using the JUNG [12] data model. The engine employs the concept of a worker pool, wherein every discrete time instance in the trace is assigned to a specific worker that builds its respective connectivity graph. Depending on the hardware capabilities (i.e. number of cores) of the comptuting system where VEGA is running, the build engine might spawn several concurrent workers for speedup purposes. Once a worker terminates the build process, it takes the next available time instance in the queue.

Mobility Trace Importer: This module is responsible for importing various formats of raw mobility traces, either obtained from real or synthetic datasets. Currently, VEGA supports importing the following mobility traces: (i) ns-2 and ns-3, (ii) SUMO Floating Car Data (FCD), (iii) VanetMobiSim, (iv) Shanghai real traces, (v) CabSpotting real traces. The user selects the type of the trace format through the configuration file. Coordinates of the trace file are converted to the Euclidian space dimension, regardless of the native coordinate system.

Road Map Importer: This module is responsible for importing various types of road network maps. It takes representations of road networks in different formats, and converts them in an internal representation that can be utilized either by the Graph Builder engine, or the Graph Analysis component. Currently, VEGA supports importing road maps from the following mapping representations: (i) OpenStreetMap.org [14], (ii) U.S Census Bureau TIGER maps [27], and (iii) SUMO mobility simulator maps. If available, the module can be instructed and configured to enrich each road network with historic traffic statistics. Nevertheless, more mapping services can be supported, simply by implementing a set of interfaces.

Graph Exporter: VEGA allows the export of V2X in a number of industry-standard graph formats. Currently, the following formats are supported for export: GEXF, Pajek, GML, GraphML [4, 6, 8, 15] and Matlab's sparse graph. This enables other 3rd party frameworks like Gephi, NetworkX, R to import V2X graphs generated by VEGA for further processing and analysis.

Graph Storage: VEGA supports peristent graph storage to 2 datastores, particularly to DBMSs or the local file system. For DBMS, VEGA has connectors to MySQL and Postgresql systems. Regarding the local file system, graphs are serialized sequentially into binary files with indexing based on the time instance. The datastore of preference is simply selected via the respective configuration file.

Graph Analysis: VEGA builds upong the JUNG framework and supports the following graph analyses: network size (nodes and edges), node degree, degree distributions, network diameter, average degree of separation, path redundancy, bi-connected components, triangle detection, network connectivity, number of components, global component co-efficient, node clustering co-efficient, node centrality (betweeness, closeness, information, lobby index), number of communities, community size distribution.

7.3 Large-Scale Vehicular Network Simulations

For the purposes of establishing large-scale vehicular network simulations the ns-3 [23] network simulator was utilized. However, at the time of using the platform, native support of the vehicular environment was limited. To this end, we extended the simulation platform with additional modules and VANET-based routing protocols. The aforementioned modules were developed together under the *V-Sense* ns-3 package (see Figure 7.10) and this section provides an overview of the work contributed towards this goal.

StreetMap Module Aims in providing nodes in the simulation, navigable information regarding features of an underlying road topology such as junctions and roads. This type of information includes the speed limits, number of lanes and direction of any road, the number of incoming/outgoing roads from a given junction, and presence or not of traffic lights at junctions. The street map information us stored in a graph-like data structure in order to allow easy and fast navigability using various graph algorithms such as Breadth-First or Depth First search, Dijkstras shortest path, Yens Top-K paths, etc. This includes the following components:

- *Street*: An abstraction of a road. It includes methods for obtaining its name, the junctions it connects, its length, maximum speed limit, number of lanes,

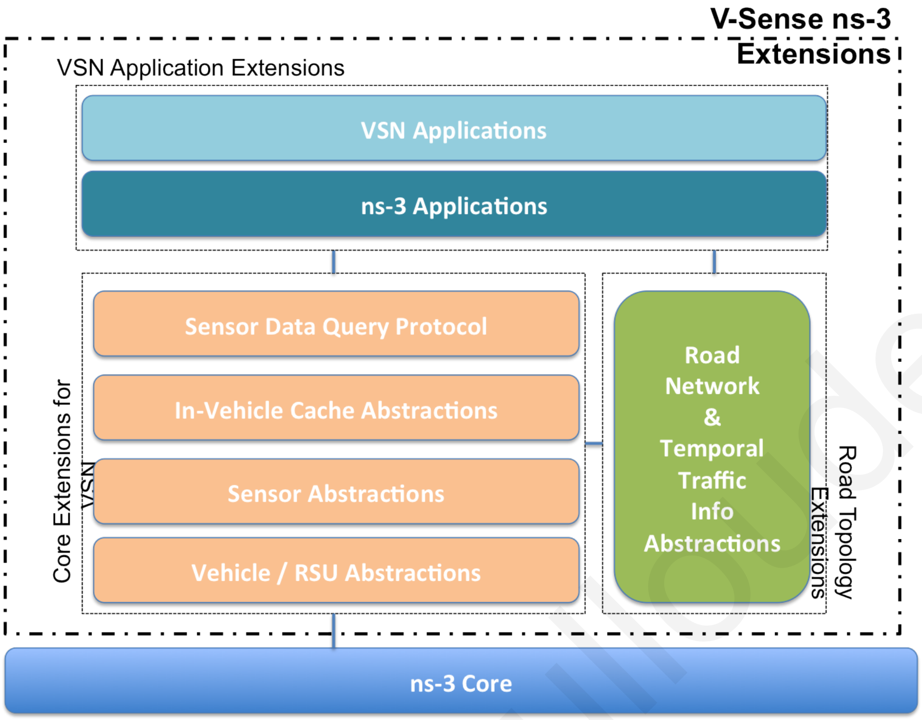


Figure 7.10: The V-Sense ns-3 Package Architecture

one-way, bearing. It also provides methods for obtaining different historical statistics if these are provided.

- *Junction*: An abstraction of a junction. It includes methods for obtaining its geographic position, its name (if it exists), the presence of traffic lights (signalized or not) and its unique id (uid).
- *StreetMap*: A graph-like data structure for representing a road network topology. It provides methods for doing several operations: (1) Add/Remove Street or Junctions (2) Find a street or junction using their uids or names (3) Get incoming or outgoing streets of a given junction (4) Get street between two junctions (5) Calculate weighted Dijkstras shortest path between two arbitrary junctions (weight street length).
- *StreetStatistics*: Holds traffic statistics for a particular street. This statistics are average speed, vehicle density, arrival and departure rates.

- *Osm-StreetMap-Helper*: Helper for parsing OpenStreetMap (http://www.openstreetmap.org) files. The parsers read the road network structure and map feature information from the XML content in the OSM map file. Then it populates the StreetMap graph structure. It also installs the given OSM map to all nodes in the simulation. Parser is for OSM network format version 0.6 and later.

- *SUMO-StreetMap-Helper*: Helper for parsing SUMO network files. The parser reads the road network structure and map feature information from the XML content in the SUMO map file. Then it populates the StreetMap graph structure. It also installs the given SUMO map to all nodes in the simulation. Parser is for SUMO network format version 0.13 and later.

- *Sumo-StreetStatistics-Heper*: Helper for parsing SUMO network dumps containing network statistics. The parser reads the traffic statistics from the XML content in the SUMO network dump file and populates the StreetStatistics component for each road in the topology. Parser is for the SUMO network format version 0.13 and later.

Vehicle Assisted Data Delivery (VADD) Routing Protocol Module - Has been implemented in order to model the VADD routing protocol [126] for ns-3. The specific module implements all required components that enable the correct function of the protocol as per the original scientific article proposing it. It it worth noting that this is the first implementation of the VADD protocol for ns-3 (and first publicly available VADD source code), which can serve as an example for implementations of other VANET-based protocols. It is also the first publicly available VADD implementation [5]. It includes the following components:

- *Vadd-Routing-Protocol*: responsible for orchestrating the overall functionality of the protocol and instantiating the module in the simulator. It provides support for all the intersection modes that are presented in the VADD paper.

- *Vadd-Packet*: holds the necessary structures that describe a packet exchanged between VADD peers. The beacon packet contains information about the position, heading and velocity that are periodically broadcasted by vehicles. VADD data packets contain information about the multi-hop forwarding of queries to a target location. These include, geo-coordinates of the vehicle, the target location, the query-lifetime (expiration), hop count, sequence number, etc. It contains methods for accessing the above packet fields and encoding/decoding a packet for transmission.

- *Vadd-Queue*: is responsible for queuing packets for transmission from a vehicle. Packets are queues on a time order. It contains methods for adding, removing and purging a queue.

- *Vadd-Matrix*: is responsible for performing the matrix operation required by the protocol to decide the best direction where packets should be forwarded in order to reach the destination with the smallest delay. It provides methods for calculating and solving a particular matrix using Gaussian elimination [56].

- *Vadd-Road-Delivery-Delay*: is used to hold calculated information about the delivery delay that will be imposed on a packet if a specific road is chosen for delivery from one junction to another. It provides methods to sort the outgoing roads of a particular intersection based on the projected delivery delay.

- *Vadd-Helper*: helper class for installing the VADD routing protocol to ns-3 nodes.

Greedy Perimeter Coordinated Routing (GPCR) Protocol Module This module has been implemented in order to model the GPCR routing protocol for ns-3. The specific module implements all required components that enable the correct function of the protocol as per the original scientific article proposing it [82]. Again this is the first implementation of the GPCR protocol for ns-3 (and first publicly available GPCR source code) [5]. It includes the following components:

- *Gpcr-Routing-Protocol*: responsible for orchestrating the overall functionality of the protocol and instantiating the module in the simulator. It provides support for all the node role detection methods proposed in the GPCR paper (Correlation co-efficient or neighbor-tables).

- *Gpcr-Packet*: holds the necessary structures that describe a packet exchanged between GPCR peers. The beacon packet contains information about the. position, heading and velocity that are periodically broadcasted by vehicles. GPCR data packets contain information about the multi-hop forwarding of queries to a target location. These include, geo-coordinates of the vehicle, the target location, the query-lifetime (expiration), hop count, sequence number, etc. It contains methods for accessing the above packet fields and encoding/decoding a packet for transmission.

- *Location-Service*: functions as a real-life location service in order to identify query the location of particular vehicles. It provides functions to retrieve the coordinates of one or more vehicles. This information are obtained from the ns-3 mobility model.

Landmark Overlays for Urban Vehicular Routing Environments (LOUVRE) Protocol Module This module has been implemented in order to model the LOUVRE routing protocol [73] for ns-3. The specific module implements all required components that enable the correct function of the protocol as per the original scientific article proposing it. This includes the following components:

- *Louvre-Routing-Protocol*: responsible for orchestrating the overall functionality of the protocol and instantiating the module in the simulator. It allows the distributed estimation of road density and thus the creation of overlay nodes in junctions where there is sufficient vehicle presence.

- *Louvre-Packet*: holds the necessary structures that describe a packet exchanged between LOUVRE peers. The beacon packet contains information about the position, heading and velocity that are periodically broadcasted by vehicles. LOUVRE data packets contain information about the multi-hop forwarding of queries to a target location. These include, geo-coordinates of the vehicle, the target location, the query-lifetime (expiration), hop count, sequence number, etc. It contains methods for accessing the above packet fields and encoding/decoding a packet for transmission.
- *Louvre-Rtable*: functions as a routing table for the overlay network.

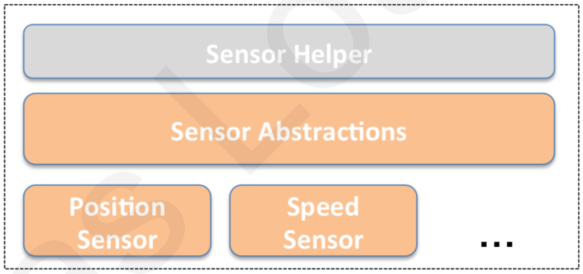


Figure 7.11: The V-Sense Sensor Abstraction

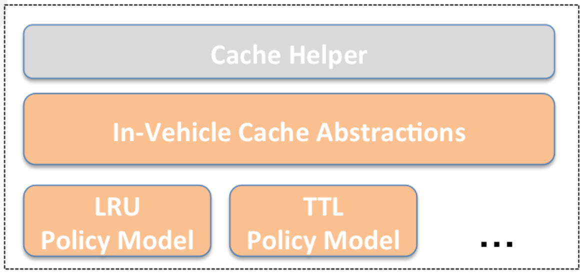


Figure 7.12: The V-Sense Cache Abstraction

V-Sense Architecture Module The V-Sense architecture was implemented in the ns-3 simulator. Particularly, the implementation provides abstractions for core actors in a VANET based environment wherein information sensing takes place. Such abstractions are those of vehicle sensors, on-board caches and cache-replacement policies, as well as data aggregation modules. easy to comprehend, extensible abstractions where any interested simulator user in the future can extend their functionality to provide specific implementations that will assist in the evaluation of VSN dedicated routing protocols/applications. The implementation of these abstractions follows the ns-3 module extension mechanism, allowing thus other simulation modules (routing protocols, applications, monitors, etc.) to easily interact with them. Furthermore, exemplary implementations of these abstractions were also developed. For instance, for the sensor abstractions (see Figure 7.11, exemplary implementation for position and speed monitoring were implemented. For in-vehicle cache abstractions (see Figure 7.12, two different information invalidation models were developed such as the Least Recently Used (LRU) and Time-To-Live (TTL). Moreover, helper modules for easily deploying and utilizing the above abstractions were developed. The V-Sense architecture implementation contains the following components:

- *Vehicle*: An abstraction of a vehicle. It is an extension of the ns-3 core node object with additional fields such as vehicle VIN number, route itinerary (if that exists and is provided).

- *Road-Side-Unit*: An abstraction of a RSU. It is also an extension of the ns-3 code node object with additional fields such as the fixed geographic position, its type (Service advertiser, Bridge), any on/off times and geographic coverage.

- *Sensor*: An abstraction of a vehicle sensor. It provides methods for setting or getting the sensor type (operational, environmental, supplementary, etc.), the information refresh interval, sensor state (running, sleep, off) and privacy. It also contains event notification registers, where someone might want to be able to be automatically informed when the sensor changes its state.

- *Speed-Sensor*: An extension of the Sensor component and specifically an exemplary implementation of a speed sensor. Such sensors record the vehicle speed (m/s) periodically and inform any listeners.

- *Sensor-Manager*: An abstraction of the in-vehicle sensor managing entity. It provides methods for getting the list of available sensors in a vehicle, the list of public sensors, to add/remove a particular sensor, or to get a handle to a particular type of sensor - if that exists. (i.e. position-sensor).

- *Sensor-Helper*: Helper for installing one or more sensors to a vehicle or a RSU.

- Cache-Eviction-Policy: Provides an implementation for the following cache item invalidation models: (1) Least Recently Used - LRU, (2) Most Recently Used - MRU, (3) Time-To-Live - TTL, and (4) First-In-First-Out (FIFO).

- *Cache*: An abstraction of the in-vehicle cache. It provides methods for setting and/or getting the maximum cache size, the current utilization, refresh interval and cache replacement policy (LRU, MRU, TTL, FIFO, etc.). It also provides an interface for invoking cache invalidation on demand.

- *Cache-Helper*: Helper for installing a cache entity to a vehicle or a RSU and associating with a particular cache eviction policy.

- *Route-Itinerary*: Holds a route itinerary for a vehicle (sequence of roads in a trip).

Chapter 8

VANET Evolutionary Dynamics Through Complex Network Science

8.1 Basic Concepts and Definitions

This section contains the definitions of the metrics used in the study. We categorize the examined metrics as *network-oriented* , *centrality*, *link duration* and *component-oriented*. All node IDs mentioned in this section refer to the sample graph of Figure 8.1. We model the VANET as an undirected graph $G(t)$, where vehicles correspond to the set of vertices $V(t) = \{u_i\}$ and communication links to the set of edges $E(t) = \{e_{ij}\}$. An edge $e_{ij}(t)$ exists, if u_i can communicate directly with u_j at time t, with $i \neq j$.

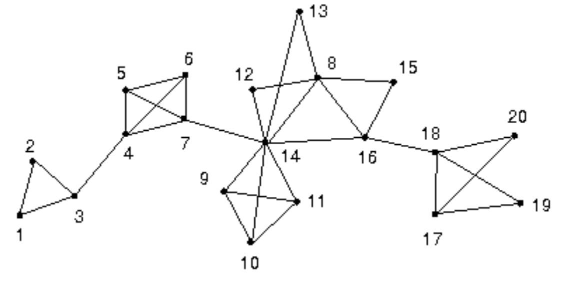


Figure 8.1: Snapshot of a sample VANET graph.

8.1.1 Network-oriented metrics

Network-oriented metrics depict the shape of VANETs, capturing and quantifying the richness of network connectivity. These metrics provide a high-level view in the topological properties of a vehicular network, stimulating initial considerations on how network protocols could be designed to take advantage of vehicular mobility to improve their performance.

- **Node degree**. The number of vehicles within the transmission range of a node. Formally, the degree of u_i at time t is defined as:

$$D_i(t) = \|\{u_j \mid \exists e_{ij}(t)\}\|. \tag{8.1}$$

- **Effective Diameter**. The minimum distance in which the 90th percentile of all connected pairs of vehicles can communicate with each other. It is a smoothed form of network diameter which we use for our studies.
- **Path Redundancy**. It is the maximum number of disjoint edges between two connected nodes in the network.
- **Network Density**. Defined as the ratio between the number of edges in the $G(t)$ and the maximum number of edges possible for $G(t)$.
- **Geographic Diameter**. Defined it as the length of the effective graph diameter considering the real underlying geographic topology. It conveys the geographic span of the shortest path between the most distant nodes in the graph and is measured in meters instead of hops.
- **Network Connectivity [116]**. Defined as the maximum fraction of vehicles that are directly or indirectly connected in $G(t)$. Formally, network connectivity is defined as:

$$NC \overset{d}{=} \max_i \left\{ \frac{1}{|V(t)|} \sum_j A(u_i, u_j) \right\} \tag{8.2}$$

where $A(u_i, u_j)$ is a connectivity indicator which takes the value of 1 if a path is available from vehicle u_i to vehicle u_j at time t, and 0 otherwise.

8.1.2 Centrality metrics

Centrality metrics have been developed within the scope of complex network science analysis to quantify how important particular individuals are in a network. Centrality metrics are designed such that the highest value indicates the most central node. Previously, in [97], we observed several centrality metrics and concluded that for VANETs, Betweenness Centrality and the Lobby index provide a good measure as they relate to the expected role a node plays within the vehicular ad hoc network. These metrics are relatively static in time (i.e traditional social networks) and thus might provide misleading conclusions in a highly dynamic setting such as the one found in a VANET. In this work, we include the results of other common centrality metrics, however we adopt also a new measure called Communicability that takes into account time dependency and its effects.

Static Network Centrality Metrics

- **Betweenness Centrality [117]**. Defined as the fraction of the shortest paths between any pair of nodes that pass through a node. The betweenness centrality of a vehicle u_i at time t is:

$$BC_i(t) = \sum_{j \neq k} \frac{sp_{jk}(u_i, t)}{sp_{jk}(t)} \tag{8.3}$$

  where sp_{jk} is the number of shortest paths linking vertices j and k at time t and $sp_{jk}(u_i, t)$ is the number of shortest paths linking vertices j and k that pass through u_i at time t. Betweenness centrality is a measure of the extent to which a vehicle has control over information flowing between others (e.g., $BC_{14} = 0.68, BC_8 = 0.07$).

- **Closeness Centrality [117]**. Defined as the inverse of the sum of the distances between a given node and all other nodes in a connected component. The closeness centrality of a vehicle u_i at time t is:

$$CC_i(t) = \left[\sum_{j=1}^{n} d(u_i, u_j) \right]^{-1} \tag{8.4}$$

where $d(u_i, u_j)$ is the distance between vehicles u_i and u_j. Closeness centrality reflects how close (in terms of distance) one node is to other nodes in the network, thus measuring its productivity when it comes to spreading/communicating information in the network. The idea is that centrality is inversely related to distance. In the example, u_{14} has $CC_{14} = 0.5$, indicating that it can spread information faster than u_{17}, with $CC_{17} = 0.26$

- **Information Centrality [112]**. Defined as the the harmonic average of the "total information" between node i and every other node j. The information centrality of a vehicle u_i at time t is:

$$IC_i(t) = \frac{N(t)}{\sum_{j=1}^{n} \frac{1}{I_{ij}}} \tag{8.5}$$

where $N(t)$ is the number of nodes in the network under examination at time t, and I_{ij} is the sum of the information in all paths between the nodes i, j. The information contained in a path between two nodes i, j in a graph is defined as the inverse of the length of the particular path. In contrast to betweeness centrality, which focuses only on geodesics, information centrality is a measure of the extent to which a vehicle has control over information flowing in all paths (not only shortest paths) between nonadjacent vehicles in the network.

Given the dynamics of the VANET, and the inability to create and maintain routing states that guarantee information flows via the available shortest paths, the information centrality thus makes an interesting metric to study. In the example, u_{14} has $IC_{14} = 0.07$, whereas u_{17} has $IC_{17} = 0.03$

- **Lobby Index** [69]. The lobby index of a given vehicle u_i at time t, denoted as $L_i(t)$, is the largest integer k such that the number of one-hop neighbors of u_i in graph $G(t)$ with degree at least k equals k. This metric can be seen as a generalization of a node's degree at t, conveying information about the neighbors of the node as well (e.g., $LI_8 = 3$ and $LI_{15} = 2$).

8.1.3 Link level metrics

The metrics of this category reflect the network connectivity over a period of time. These metrics are critical both for the periodic single-hop broadcast primitive, where a vehicle broadcasts its current information (location, velocity) in its direct neighborhood, and also for the multi-hop geocast primitive, where several data exchanges between vehicles take place during one communication session.

- **Number of connected periods**. The number of established links between a pair of vehicles within a given time period. A connected period is the continuous time interval during which a physical link is established between two vehicles as a consequence of one being in the transmission range of the other.

- **Link duration**. The time duration of a connected period. Formally, the duration $l_{ij}(t)$ of the link from u_i to u_j at time t is defined as $l_{ij}(t) = t_c - t_o$, if $\exists e_{ij}(t)$, where $t \in [t_o, t_c]$ and $\nexists e_{ij}(t')$, where $t' < t_o$ or $t' > t_c$.

- **Re-healing time**. The time span between two successive connected periods of a pair of vehicles.

8.1.4 Component-oriented metrics

The examined metrics present the dynamic properties of components and dense sub-graphs established inside a VANET communication graph. The existence of vehicle communities is important in terms of information propagation, since they can act as "data islands" in a vehicular environment.

- **Number of Components**. The number of co-existing, non-connected clusters of nodes at a given instant. We define as component a connected group of vehicles. A connected group is a sub-graph of the network such that there is a path between any pair of nodes.
- **Component Coefficient**. It measures the degree to which nodes in a network tend to cluster together.

  The local component coefficient $LCCF_i(t)$ of a node i at time t (as defined in [118]) quantifies how close the neighbors of i are to being a clique and is defined as:

  $$LCCF_i(t) = \frac{2|E_k(t)|}{|N^k(t)|(|N^k(t)| - 1)}, \tag{8.6}$$

  where $|E_k(t)|$ is the number of existing links in cluster k at time t and $|N^k(t)|$ is the number of nodes in cluster k at time t. $LCCF_i(t)$ has a maximum value 1 if the neighbors of i are a clique and 0 otherwise.

  Accordingly, the global component coefficient $GCCF_k(t)$ of a component k at time t is:

  $$GCCF_k(t) = \frac{\sum_{i \in k}(LCCF_i(t))}{|N^k(t)|} \tag{8.7}$$

- **Number of Communities [96]**. The number of existing communities at a given instant. A community is defined as a dense sub-graph where the number of intra-community edges is larger than the number of inter-community edges.

In order to identify communities, we transform $G(t)$ to directed graph so as $D_i^{in}(t) = D_i^{out}(t) = D_i(t)$, where $D_i^{in}(t), D_i^{out}(t)$ is the in-degree and out-degree of node i at time t. Formally, a sub-graph $U(t)$ of a VANET graph $G(t)$ at time t constitutes a community, if it satisfies:

$$\sum_{i \in U}(D_i^{in}(t))(U(t)) > \sum_{i \in U}(D_i^{out}(t))(U(t)), \tag{8.8}$$

i.e., the sum of all connections within the community $U(t)$ is larger than the sum of all connection toward the rest of graph the $G(t)$.

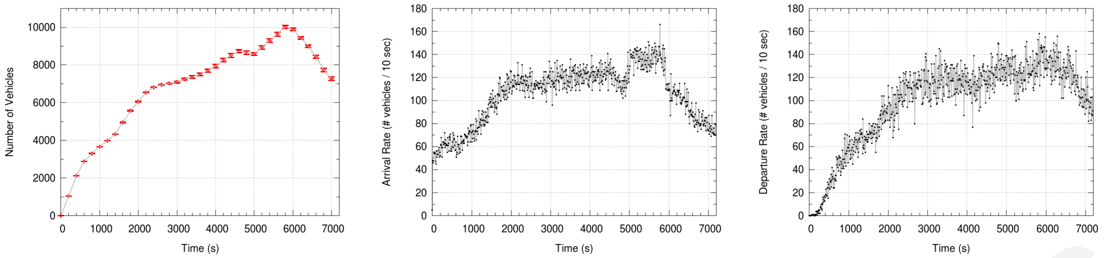
- **Community Modularity [96]**. Community modularity quantifies in the range of -1 to 1 the division of the graph into communities. Good divisions, which have high modularity values, give communities with dense internal connections and weak connections between different communities. The community modularity Q of a network at time t is:

$$Q_t = \frac{1}{2m}\sum_{ij}[A_{ij} - \frac{D_i D_j}{2m}]\delta(i,j) \tag{8.9}$$

  where A_ij is an element of the adjacency matrix of the network, m number of links and D_i the degree of node i.

8.2 Analysis TestBed

This section provides an in-depth overview of the dataset used in the study of the VANET communication graph, as well as the analysis testbed setup.



(a) (b) (c)

Figure 8.2: Vehicular Traffic Statistics: (a) Number of Vehicles, (b) Arrival Rate and (c) Departure Rate. (Red bars indicate the standard error between 10 different simulator runs - Plotted every 100s)

8.2.1 Mobility Insights

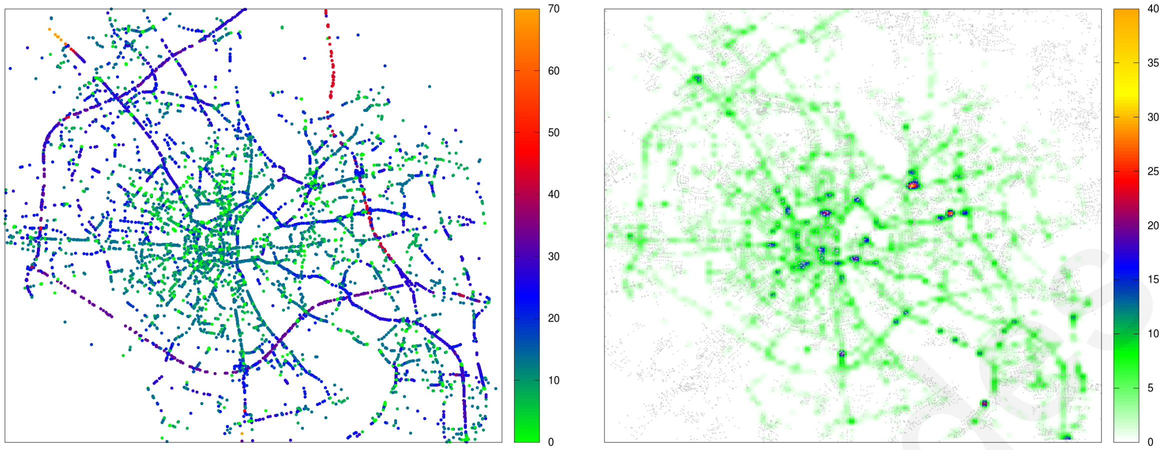
In this work the TAPAS-Cologne dataset was utilized (see Section 7). For the purpose of our study, we focus on the morning traffic rush-period between 6:00am and 8:00am, taking snapshots of the network every 1 second. Here we give insights on the characteristics of vehicle mobility in this dataset.

During the 7200 second time window, approximately 75600 unique vehicles enter and travel within the aforementioned area.[1] Figure. 8.2, depicts the total number of vehicles in the area of Cologne every 10 seconds, in conjunction with the number of new arrivals and departures at each time snapshot. The diagram plots the average values (and respective standard error) over 10 different simulation runs, each one with a different random seed generator (see Sub-Section 8.4 for more info on the simulation runs).

We observe, that the number of vehicles concurrently traveling in the area increases with time with a phase-shift around 7:35am-7:40am, pertaining thus to the characteristics of the morning rush hour that are traditionally encountered in large cities. Arrival and departure rates, present significant fluctuations in between consecutive time instances during the complete 2-hour time window of our study.

---

[1] Similar traffic patterns are also encountered during the afternoon period between 5:00pm-7:00pm where people return back home from work.



(a)

(b)

Figure 8.3: (a) Average Vehicle Speed (Km/h), (b) Vehicle density (v/200m^2)

Figure 8.3(a), presents a snapshot of the traffic status at 6:30 am ($\tau = 1800s$), with each vehicle colored based on its instantaneous speed. By referencing the road topology map from Figure 7.3, we observe that vehicles with significantly higher speeds (blue color) can be found travelling on high flow freeways and arterial roads that form concentric circles around the city center. As expected, lower speed and congested traffic (green color) is found within the city center, mainly travelling on urban and collector roads. Due to the presence of a large number of activity places such as work offices and shopping stores, as well as the existence of several intersections that interrupt the normal flow of traffic, as it can be seen from Figure 8.3(b) vehicular traffic is considerably more dense in urban roads. By examining the mobility dataset even further, we identify that in the 2-hour time frame we study, the majority of vehicles have an average speed (in terms of the whole trip duration), which is less than or equal to 45 Km/h (12m/s). Also, most vehicles have a total travel time which is less than or equal to 14 minutes. These findings are in accordance to the curvemetric distance[2] of the 90th percentile of vehicles, which is less than or equal to 15Km. In turn, the above indicate that the majority of vehicles in the dataset take short trips

---

[2] The distance measured between a vehicles' starting and destination point, following the geometric shape of the underlying road topology.

(both in terms of time and distance), and these trips traverse some part of the dense inner city center.

8.3 Wireless Communication Technologies

Another key factor that undoubtedly determines the structure and evolution of the VANET communication graph is the technology through which vehicle-to-vehicle (V2V) wireless communication is achieved. Currently, the de-facto wireless communication protocol for V2V is designated as IEEE 802.11p. Is an amendment to the IEEE 802.11 standard has been proposed to add wireless access in vehicular environments (WAVE). It defines the necessary enhancements to support data exchange between high-speed vehicles and between vehicles and roadside infrastructure. The IEEE 802.11p protocol is still under development and only limited trial implementations exist. To the best of our knowledge, the only available field measurements for the IEEE 802.11p are the communication performance results in [91] from the CVIS project.[3] These results indicate that, irrespective of vehicle speed, effective communication can be achieved at a distance of 300m with a good and relative constant data rate (5 Mbit/s). Our literature review, has indicated that the value of 300m is also referenced in [124] as the indicative V2V communication distance using the IEEE 802.11p protocol. Nevertheless, as also stated in [116], a constant unit disc transmission coverage (i.e 300m) is far but realistic in urban environments, due to the effects that buildings and other obstructions might have in the propagation and strength of wireless signals. To this end, we assume the model of line-of-sight(LOS) and non-line of sight (NLOS) wireless communication, wherein vehicles establish a wireless link if they are within 250 and 140 meters of each other, respectively. For

---

[3] CVIS technology is using the CALM M5 ISO standard that incorporates the IEEE 802.11p PHY/MAC

more information on when two vehicles might have LOS or NLOS communication, the reader is referred to [116].

8.4 Testbed Overview

Location	**Cologne, Germany**
Trace Type	Realistic
Area Size (Km)	33 × 35
Average Vehicle Speed (m/s)	12.6
Total Vehicles	75570
Vehicle Transmission Range	LOS:250m, NLOS:140m (802.11p)
Simulation Time	7200 sec (1000s warm-up)
Snapshots Interval	every 1 second

Table 8.1: VANET Evolutionary Dynamics - Testbed Parameters

We provide the necessary input files (road map, route definitions, etc) obtained from [115] to the SUMO simulation framework and instruct it to run 10 times, each one with a different random seed generator. We allow a 1000 second warm-up period at the beginning before obtaining any measurements in order to achieve some level of stability in the network. Consequently, we study snapshots of the VANET communication graph taken every 1 second. Overall, approximately 436000 snapshots of the VANET communication graph were recorded. Table 8.1 provides an overview of the evaluation testbed.

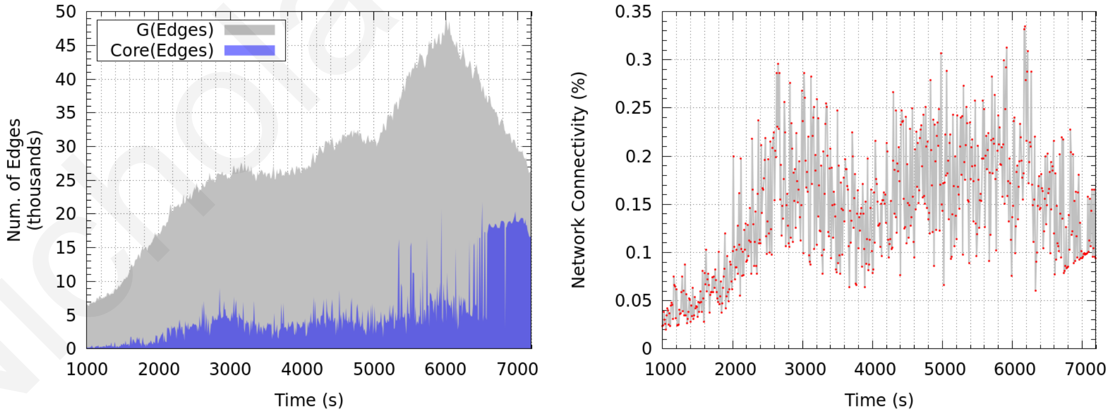
8.5 Observations

This section presents the findings of our study related to the laws governing the networking shape of vehicular connectivity by considering the realistic testbed presented previously. Specifically, we drive our analysis, by posing a series of questions that aim to extract a “higher order” knowledge of the time-evolving topological characteristics of the VANET communication graph and at the same time present the effects that mobility has on such networks.

Note that our work does not examine the dynamics of the VANET communication graph under the presence of medium access (MAC) problems such as contention and interference [64]. Our work shows that, based on factors such as: (i) the mobility pattern of vehicles and (ii) a given transmission model, there is a possibility that a number of communication links could be established (if required) at a given point in time. The set of all such possible links, determine the topology and characteristics of the VANET graph, and knowledge of such information is crucial in the design/-operation of routing and data dissemination protocols. However, the particulars of how these links (i.e the shared medium) are accessed, is the work of MAC protocols, something which is out-of-scope of this work.

The data analysis was performed using the VEGA framework, that is described in Section 7.

Question #1: What is the size of the VANET and how does it behave in the arrow of time?



(a) (b)

Figure 8.4: (a) Size of the VANET and its core, (b) Network Connectivity, across time.

We refer to the size of the VANET as the number of edges (i.e. potential wireless connections) that are present in the communication graph at any given time instance. An edge in the communication graph is formed if two vehicles concurrently moving on the underlying road topology, are within the LOS/NLOS transmission range of each other. As the grey-shaded plot in Figure 8.4(a) depicts, the size of the VANET exhibits a constant and smooth growth trend across the time arrow. This trend follows closely the one observed when exploring the number of vehicles that enter the area of Cologne (see Figure 8.2). Conversely, the VANET size is not sensitive to the high variability exhibited by both the arrival and departure rate of vehicles. For instance, although we have observed that in several time instances an increase in the number of departing vehicles can cause a decrease of the VANET size, this effect is mitigated by the set of vehicles still in the network, whom their collective mobility profile can cause the sustainability of current wireless links and even the establishment of new ones. Interestingly, the VANET core[4] geography, is very well embedded (overlaps) to the underlying geography of the city center. Averaging across time, the core includes 18% of the total edges (blue-shaded area on Figure 8.4(a)) and 7% nodes available in the network, respectively.

Question #2: How well is the VANET connected across time?

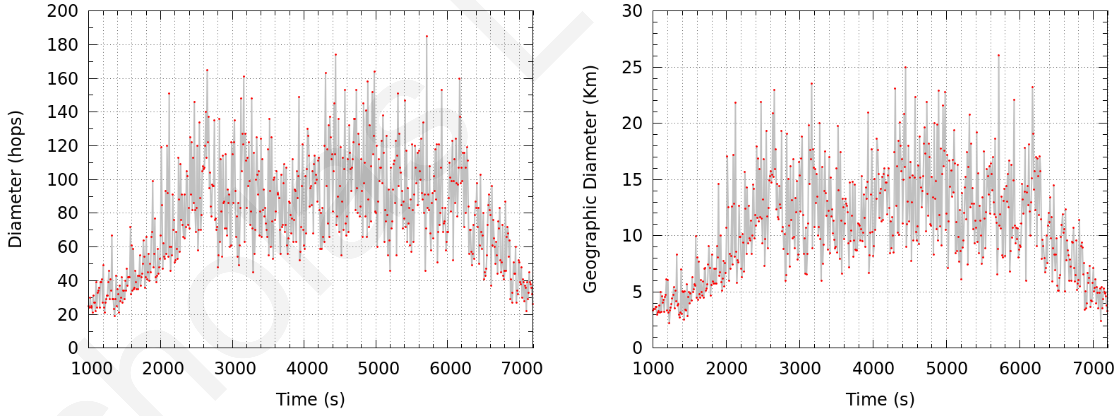
Initial insights on the global connectivity conditions that prevail in the VANET, report that this is a very sparse network. Even in the case of our study, where we consider 100% V2V technology penetration ratio and high vehicle density (due to rush-period), the corresponding network density is very low and never exceeds 0.4%. Figure. 8.4(b) plots the global network connectivity (Equation 8.2) for the 6200s we analyze. *NC* is upper bounded roughly at 35% of the total number of participants that exist concurrently in the geographic area under study. In other words, in an optimal global connectivity scenario, only 35% of the total vehicles in the topology can

---

[4] The core is defined here as the largest biconnected component existing in the network.

exchange information either using a direct or a multi-hop communication paradigm. The nature of NC provides a birds-eye view of the VANET capacity to facilitate direct or multi-path inter-connectivity for its member vehicles, in the presence of various constrains induced by a large-scale urban environment (morphology of road topology, inter-vehicle distances, mobility patterns etc). These findings serve as a precursor of the VANET fragmentation level and essentially indicate the necessity of utilizing supporting mechanisms for city-wide, multi-hop and geocasting communication paradigms. For instance, the installation of fixed roadside infrastructure (V2I) can improve link stability and also serve as gateway to long range V2V information exchange. In addition, intelligent caching solutions can be an option given that the application requirements permit it.

Question #3: What is the diameter of the VANET and how does it change over time?



(a) (b)

Figure 8.5: (a) VANET Diameter in hops, (b) Geographic Diameter in kilometers, vs Time.

From the viewpoint of complex networks, the diameter of the network is considered as the longest shortest path between *any* two nodes in a graph. In the case of the VANET communication graph, the diameter provides an insight on the largest number of vehicles, a message would need to traverse in order to be disseminated between any connected (directly/multi-hop) vehicle pair. Since a hop in the VANET graph is primarily dependant on the geographic distance between the two connected vehicles, inevitably the largest shortest path has also a geographic dimension. Hence, we also study the geographic diameter. Figure 8.5 plots both VANET diameters, network (left) and geographic (right), against time. Evidently, the diameter of the VANET communication graph, as well as its geographic counterpart, exhibit significant variability in the arrow of time. Interestingly, we have identified several time instances where the diameter can fluctuate upwards or downwards in excess of 50 hops ($\sim 15Km$) within just a few seconds, denoting a severe or complete change in the structure of the largest shortest path in the VANET.

Question #4: Which factors affect the VANET diameter?

Given this diameter volatility, it is crucial to identify the factors or circumstances that can ultimately trigger such behavior. Specifically, what causes the partial, or total re-wiring of the largest shortest path in a matter of few seconds? Is this a side-effect of microscopic vehicle features such as geographic location, or macroscopic traffic patterns such as vehicle density, arrival and departure rates?

To provide answers to the above questions, we narrow down our analysis to a smaller time window, which includes such abrupt transitions.[5] Specifically, we focus to a 180-second window, from 6:35 a.m ($\tau = 2100s$) to 6:38 a.m ($\tau = 2280s$). We start by examining the macroscopic aspect as displayed in Figure 8.6, where the VANET diameter is plotted against the VANET size, as well as the arrival and departure rate of vehicles in the area. In real world networks that evolve over time, the graph diameter is known to be a function of network size [78], however this norm seems not to hold true for the case of the VANET communication graph. Looking at the 10-second interval between $\tau = 2170s$ and $\tau = 2180s$, we observe an increase of 89 hops and 15Km in the VANET diameter and geographic diameter, respectively.

---

[5] The specific time window was randomly chosen, however such phenomena are present in several other instances as well.

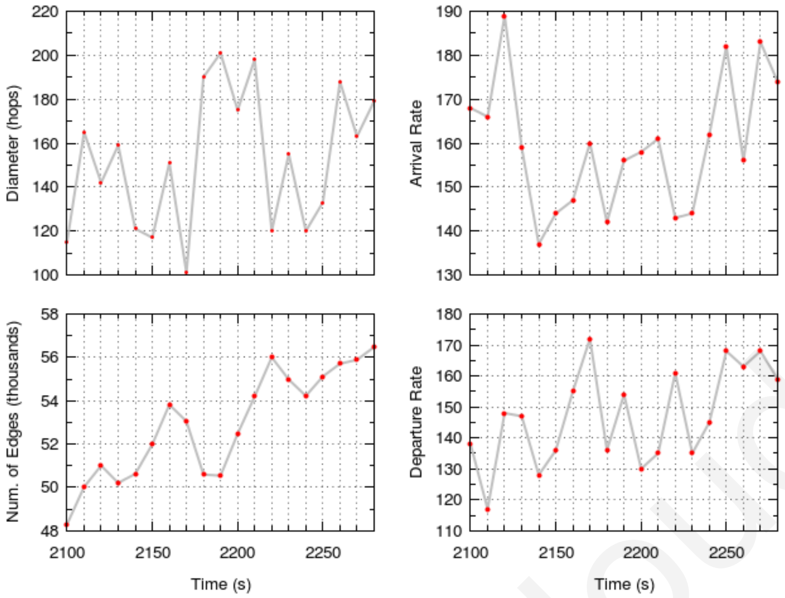
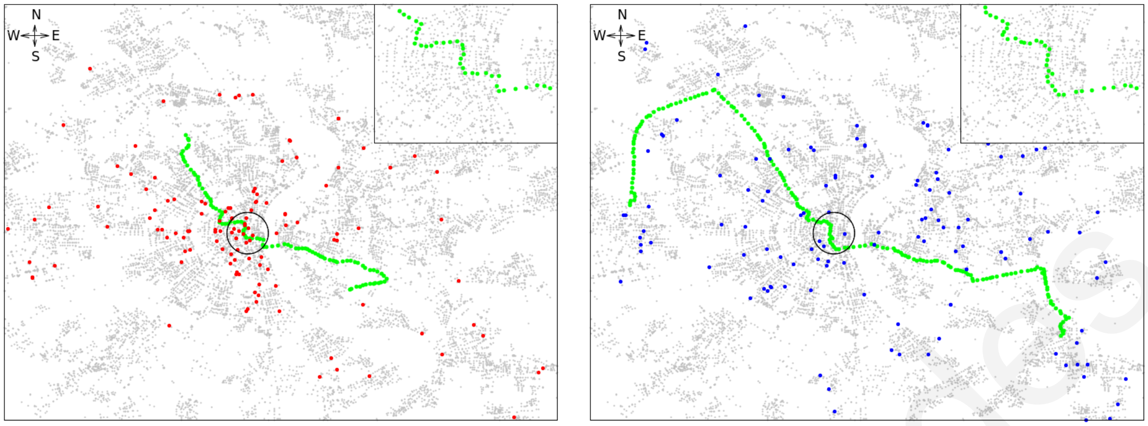


Figure 8.6: Zoom-in to time window between 6:35 a.m ($\tau = 2100s$) and 6:38 a.m ($\tau = 2280s$) (Left) Diameter and VANET size vs Time. (Right) Vehicle Arrival and departure rate (#veh/10sec) vs Time.

Intuitively, one would expect that this significant increase in the diameter would come as a result of an equally significant increase in the VANET size; that is, with more links in the network, the higher the probability to have an extended communication path (hop-wise) between a random node-pair. However, at the above time instance this expectation falls short and breaks the norm, due to the fact that the VANET size is reduced by 2454 edges.

Given the above observations we are thus inclined to engage in the following argument: Despite the reduction in the number of links in the network, a significantly larger shortest path appears at $\tau = 2180s$ which was not present 10 seconds earlier. Given that the differences in the vehicle arrival/departure rates cannot justify the respective shift in the VANET diameter (see Figure 8.6), is it possible that the path appearing at $\tau = 2180s$ comes as a result of microscopic vehicle features such as their geographic location? In other words, is it possible that these abrupt transitions of the VANET diameter, can be primarily attributed to changes in vehicles' positions?

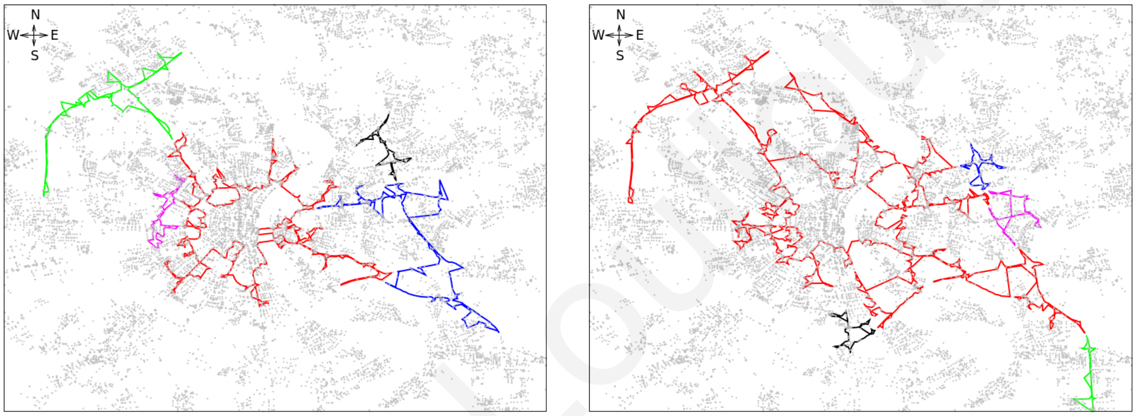


(a) $\tau = 2170s$ (b) $\tau = 2180s$

Figure 8.7: Geographic dimension of VANET Diameter overlayed over the road network. (a) Red denotes the last position of departing vehicles. (b) Blue denotes the position of newly arrived vehicles

To examine our argument, we study the position of the departing and arriving vehicles, during this abrupt diameter transition. Red color nodes in Figure 8.7(a) depict those vehicles that will depart the network at $\tau = 2170s$. Respectively, blue color nodes in Figure 8.7(b) depict those vehicles that will enter the network at $\tau = 2180s$. At first sight, one can easily observe that the position of several vehicles at $\tau = 2170s$ is such that their departure can have an adverse effect on the structure (and length) of the largest shortest path. Specifically, a subset of the departing vehicles can be seen as structural components of the largest shortest path (circled nodes in Figure 8.7(a)), and undoubtedly their removal can cause a fracture in the path itself. Indeed, at $\tau = 2180s$ we observe the segment within the circled area being rewired via a different road path (see inset plots), yet retaining its geographic proximity to the previous layout. On one hand this rewiring indicates that there still exist sufficient wireless links in the vicinity to allow the repair of connectivity gaps, while on the other hand they convey the message that departing vehicles alone are not sufficient to cause heavy alterations in the VANET diameter structure.

Moreover, we observe a significant and simultaneous extension of the VANET diameter both towards the north-west and south-east regions of Cologne. Evidently, these extensions cannot be attributed to the set of arriving vehicles since only a small fraction of those actually coincide with the newly formed largest shortest path. Our analysis has identified that these extensions are caused by changes in the internal structure of clusters present in the network, particularly the fusion of smaller clusters with the largest cluster in terms of membership.



(a) $\tau = 2170s$

(b) $\tau = 2180s$

Figure 8.8: Spatial representation of the Top-5 components (membership size) present in the VANET. Each color denotes a different component.

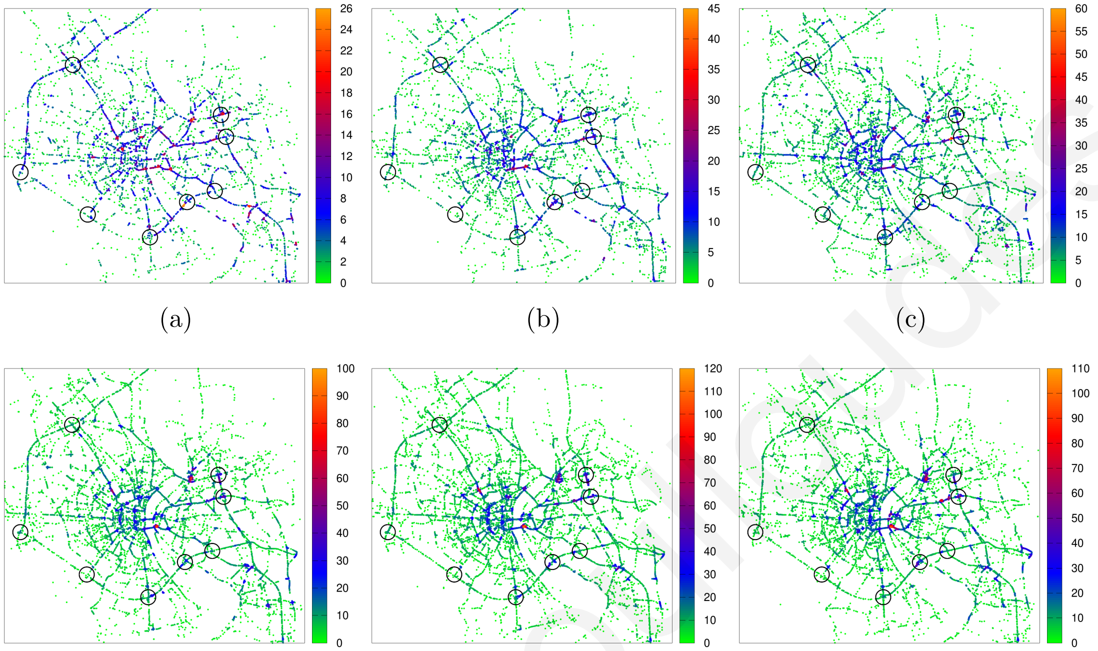
Figure 8.8 presents the shape (concave hull) of the Top-5 clusters in the VANET by considering their size. As we can see at $\tau = 2180s$, the green, magenta and blue-colored clusters that were previously present at $\tau = 2170s$, are now merged into one very large cluster (red color) that has an extensive geographic reach. Thereby, we can conclude that differences in the VANET diameter are primarily the combined effects of individual vehicle mobility, where in several nodes unawarely function as bridges among distinct components and ultimately have a strong impact on the global network topology.

On top of the previous findings, it is important to note that this variability of the shortest path is not an isolated phenomenon, rather it can be observed also in the Top-5 shortest paths (in terms of length). This indicates that the structure of communication paths in the VANET (not only for geodesics necessarily), is highly volatile and eventually can cause the degradation or even hindering of multi-hop data dissemination.

Question #5: What is the average Degree of Separation? Does it exhibit Small-World properties?

Similar to the network diameter, the average degree of separation among vehicle-pairs in the VANET exhibits high variability across time, with values ranging roughly between 8 and 67 hops. Moreover, there is significant variance between the separation of different pairs within a given time instance, indicating the lack of uniformity in the length of the existing communication paths in the network. Additionally, the average degree of separation does not change proportionally to $logN$, where N is the number of nodes in the network, thereby concluding that *large-scale, urban VANETs do not satisfy the feature and properties of small-world networks.*

Question #6: Is node degree dependant on the underlying road topology?



(a) (b) (c)

(d) (e) (f)

Figure 8.9: Spatial representation of vehicle location colored according to node degree. From upper-left: $\tau = 1000 - 3500s$ in 500s increments. Circles depict the location of large overpasses in the road network.

The diagrams in Figure 8.9 exhibit time-separated snapshots of vehicle positions while being driven on the road network of Cologne. Each vehicle is colored according to its individual node degree in the VANET communication graph. Circles denote underpasses/overpasses or exit ramps that enable vehicles transition to higher/lower grade roadways. As the day progresses and even more people commute towards the city center, there is a higher probability to establish V2V links, thereby the average node degree increases. Likewise to [93], our analysis demonstrates that the spatial distribution of node degree is not uniform throughout the topology. The VANET can be characterized as a very heterogeneous network, where isolated nodes (no neighbors), and nodes with over 100 neighbors that resemble to traditional "hubs"[6], can coexist in unity. Specifically, our analysis indicates that node degree is dependent on the proximity of vehicles to different features of the underlying road network. The majority of vehicles with lower degree (green to light blue color) can be encountered on freeways and collector roads. Multi-lane, high capacity freeways, enable neighbouring vehicles with the same direction and relative velocities to travel in platoons for a prolonged period of time. This mobility pattern fosters the establishment of a number of connections (on average between 13 and 17) among vehicles driving in the same or adjacent lanes towards the direction of travel, resulting to the degree distribution visible in the outer parts of the city. On the other hand, the reduced 1-hop connectivity on collector roads comes as a result of the low-to-moderate capacity of the roadway itself and also the presence of intersections, round-abouts and stop-signs which interrupt the continuous traffic flow.

Nevertheless, the degree of vehicles travelling on high capacity roadways is more stable over time, due to the fact that such connections are long-lived - vehicles tend to use freeways and arterial roads to cover large distances, thereby committing to longer travel times. An exception to the above occurs whenever vehicles approach various overpasses or exit ramps. In the case of Cologne, such high density road features are responsible for causing spikes in the temporal distribution of degree and in certain occasions some vehicles are able to establish a connection with up to 122 neighbors. However, it is extremely likely that these highly connected vehicles will return to a quasi-isolated state shortly after traversing the overpass or exit ramp.

Moving towards to the city center and particularly on urban roads, the node degree tends to increase substantially (darker green - blue) primarily due to the increased traffic density and lower speed limits. These factors enable vehicles to connect with

---

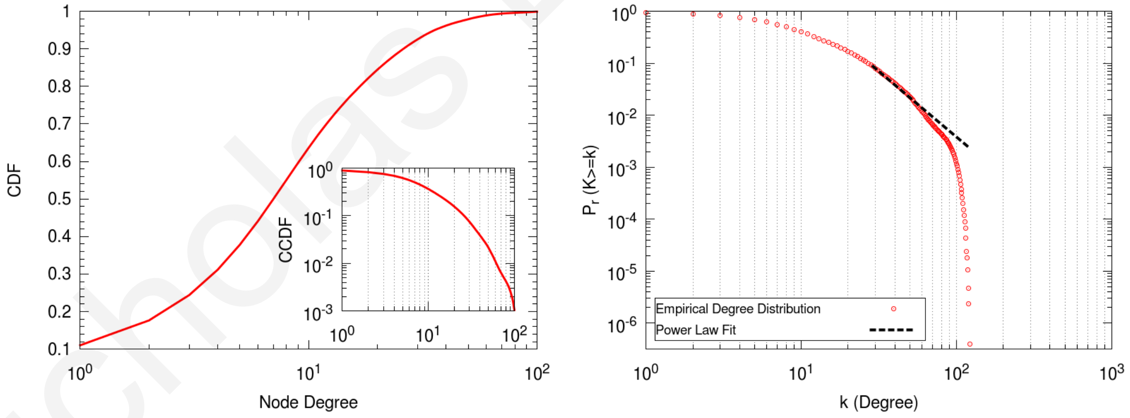
[6] In traditional complex network theory, hubs are defined as those nodes with degree equal or larger to an order of magnitude higher than the average node degree. However, in the case of the Cologne VANET, vehicles with such a degree magnitude were not identified.

Road Type	Min	Max	μ	σ
Freeway	0	101	13.7	34.0
Arterial	0	122	17.0	34.0
Collector	0	114	9.9	22
Urban	0	115	8.1	19

Table 8.2: Aggregate Vehicle Degree statistics for the different road network features.

an increased number of other vehicles being driven in the same or opposite directions, or even on various intersecting roads. However, the presence of intersections and the dense road topology with rapid transition in-between LOS/NLOS communication, makes these connections short-lived. Consequently, the degree of vehicles in these roads across time exhibits high variability. Table 8.2 presents a summary of vehicle degree statistics for each of the road features. These statistics are aggregated for the whole 6200s of our analysis.

Question #6: Does the node degree exhibit the properties of scale-free networks?



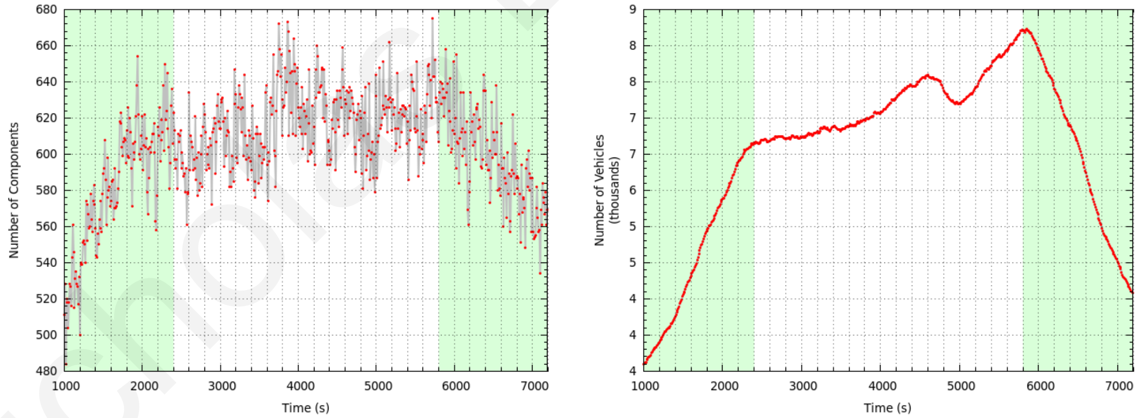
(a) (b)

Figure 8.10: (a) Cumulative and Complementary Distribution Functions of Node Degree, (b) Power-Law fit.

Figure 8.10(a) presents, the Cumulative Distribution Function (CDF) and the Complementary Cumulative Distribution Function (CCDF) of the degree distribution of vehicles in the VANET communication graph, aggregated over all snapshots in the

6200s interval we study. 50% of the vehicles have a degree of approximately 10 or less, while 90% of the vehicles have a degree of 25 or less. The degree distribution plot is in accordance with our previous findings, indicating the heterogeneity of the VANET which can simultaneously accommodate isolated nodes, and nodes with 120 neighbors. Examining the degree distribution even further, we make the observation that it does not follow a power-law for any value of the penetration ratio, consequently making *the VANET not scale-free.*[7] Counter-intuitively, we would expect to observe some closeness to scale-free networks, under the assumption that vehicles in the proximity of a densely populated intersection, could serve as hubs due to the benefits of the extended range of LOS communication in all directions. In addition, such hubs could mediate large path lengths between other node pairs. Indicatively, these findings are depicted by Figure 8.10(b), which plots (on log-log axes) the vehicle degree data (aggregated over all the time period) and a power-law distribution fit.

Question #7: What is the fragmentation level of the VANET?



(a)

(b)

Figure 8.11: (a) Number of Components, (b) Number of Vehicles, vs Time. Isolated vehicles not considered as components.

When discussing the fragmentation level of the VANET, we refer to the number of weakly-connected components that are concurrently present in the network.

---

[7] The Power-law Distributions fitting were calculated using the tools developed by A. Clauset, C.R. Shalizi, M.E.J. Newman - http://tuvalu.santafe.edu/ aaronc/powerlaws/

Initially, the number of components, provides a high-level view of how many vehicles can exchange information with each other, either using direct or indirect links. According to Figure 8.11(a), the VANET exhibits noteworthy fragmentation, since no less than 480 individual components are present in the network at any given instance. Inevitably, this decomposition of the network into different groups introduces communication gaps and eventually renders cross-network, multi-hop communication, inadequate for data dissemination. On this account, carry-and-forward are indispensable for inter-component information exchange caching mechanisms. Nonetheless we observe 3 different temporal phases in component population evolution, that match closely the evolution of vehicle population. The 1st phase (left green shaded area) exhibits a growth trend, wherein the arrival of new vehicles ($\sim 7500 vehicles$) in the topology sparks the creation of additional small components (both temporally and spatially). Specifically, in this 1st period the majority of components are comprised of 3-4 vehicles only and cover, on average, a geographic area less than $15m^2$.

During the 2nd phase we observe a relative stability in the component population growth, where newly arrived vehicles ($\sim 7500 - 8500 vehicles$) do not create additional components, but rather join existing ones. As the day progresses and existing components become even larger in terms of membership, they are fused with other smaller components in the vicinity, extending thus their corresponding geographic coverage.

Figure 8.12 visualizes this phenomenon, by plotting the Top-100 largest components in the network for two time instances, 100s apart. Each circle denotes a distinct component, with the diameter of the circle corresponding to the size of the component in terms of vehicle population. The center of each circle is the geographic centroid of the component, computed by calculating the coordinate-wise mean of member vehicles. At $\tau = 2000s$ (right plot), the geometric size of several components present in the VANET gets larger due to an increase in their membership. Similarly, other

(a) $\tau = 1900s$ (b) $\tau = 2000s$

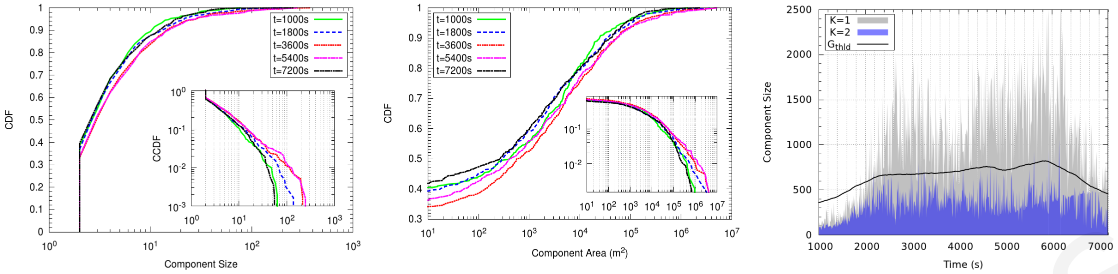
Figure 8.12: Top-100 Components Size - Circle diameter corresponds to the relative component size.

clusters that were not visible at $\tau = 1900s$ because of their small size are now easily recognizable. Interestingly, Figure 8.12, illustrates the existence of a single, very large, component in the center of the Cologne area. The particular component appears in the network 97% of the time window we are studying. As later explained, its size remains relatively the same after the above threshold vehicle density is achieved.

Finally, the phase-shift point in vehicle population ($\tau = 5800s$) triggers the 3rd phase in the component population evolution. As the rush-period moves towards the end, the resulting vehicle density is not sufficient to accommodate strong connectivity across the various existing large components, therefore causing fragmentation and reduction in the overall population.

Question #7: What is the component size distribution in the VANET?

Before capturing and examining the properties of the very large component presented earlier, it is important to gain some useful insights on the distribution of component size in the network. As mentioned, the VANET exhibits a strong fragmentation with several components being simultaneously present in the communications topology. Thereby, we would like to know how large these components can



(a) (b) (c)

Figure 8.13: Cumulative and Complementary Distribution Functions of (a) the component size, (b) geographic coverage area, (c) Top-2 Component Size vs Time.

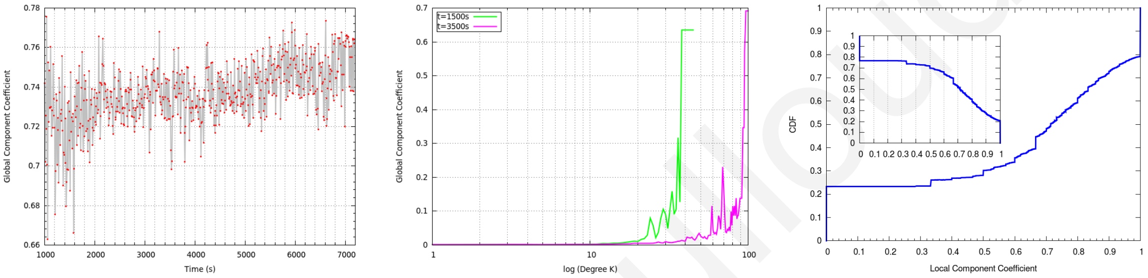
get (in terms of their membership) and ultimately what is their span in terms of the underlying geography.

Figure 8.13(a) plots the Cumulative Distribution Function (CDF) and the Complementary Cumulative Distribution Function (CCDF) of the component size, for different time instances. In contrast to [93], our work does not consider isolated vehicles (i.e singletons) in the study of VANET components, since communication of a vehicle with itself is meaningless. The VANET includes both small and large components, with 40% of those being components established among just two vehicles, while 90% of them are comprised by 15 vehicles or less. Yet, the exceedance (long tail) of the distribution depicted in the respective CCDF, conveys the simultaneous presence of certain components that are very large, and can host up to 1000 vehicles. In addition to these very large components, our study captures the existence of a single, exceptionally large component in the VANET. This component can have up to 2500 member-vehicles, albeit it appears with a lower probability. As Figure 8.13(c) depicts there always exists one very large component (K=1) which is, in fact, a Giant component since $\sim 93\%$ of the time its size surpasses the threshold $G_{thld} = (V(t) \times 0.1)$, where V is the total number of vehicles in the network at time t. In contrast, the 2nd largest component (K=2) is significantly smaller, connecting approximately up to 700 member-vehicles and can be characterized as a giant component only for a few time instances.

Furthermore, we examine how these heterogeneous components scale in terms of the geographic area they cover. Figure 8.13(b) shows the CDF and CCDF of the coverage region, calculated using the *convex hull* of each component. 37% of the components present in the VANET cover an almost negligible geographic area less than $15m^2$, conveying that the 3-4 vehicles comprising them are in very close proximity of each other. Approximately 70% of such small clusters are primarily encountered in urban roads near the outskirts of the city, where the specifics of the road-network topology (i.e. low vehicle capacity, small road-segment lengths and presence of intersections) encourage tight inter-vehicle distances. Coupled with the reduced vehicle density that characterizes such areas, the size and geographic coverage of the components formed can become very restricted. The remaining 30% of small clusters comprise vehicle platoons travelling on high speed freeways, with inter-platoon distances larger than 250 meters, which is the maximum transmission range in LOS. *Consequently, vehicles travelling in such areas, should not expect to establish and maintain communication sessions with a large number of other vehicles, either utilizing broadcast or multi-hop communication paradigms.*

Conversely, as the geographic coverage of components increases, we observe a different spatial distribution across the area of Cologne. While 90% of these geographically larger components cover an area less than $56000m^2$, now they are found with higher probability on roads close to down-town. This transition period is in line with the commute habits of people during the morning rush hour, where the majority of vehicular traffic traverses initially arterial and collector roads prior to exiting and spreading to urban roads so as to reach various activity areas (i.e workplaces, stores, amenities, etc) within the city core. The increased vehicle density encountered in these moderate-to-high capacity road structures, in conjunction with minimal travel flow interruptions, results to unhindered inter-vehicle communication that in turn facilitates the establishment of larger moving components. As vehicles travel towards the city center, their associated components follow in unison and on the way they merge with other vehicles (and/or components) eventually causing an even further growth.

Question #8: What is the structure of components in the VANET? What is the level of internal connectivity?

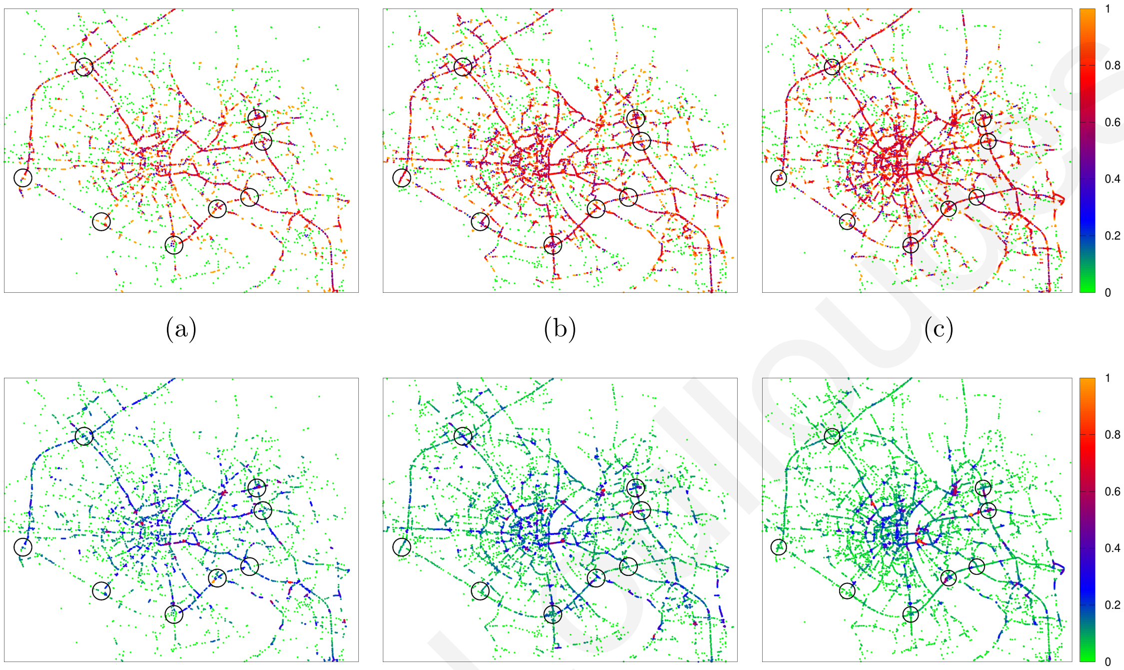


(a) (b) (c)

Figure 8.14: (a) Global Component Coefficient vs Time, (b) Global Component Coefficient averaged over all Degree K nodes, (c) CDF and CCDF of Local Component Coefficient

Now that we know that the VANET is highly partitioned into a large number of components, we would like to examine how well these components are connected internally. In other words, are these components knit in a way that can easily facilitate data dissemination to a large number of audience? The global component coefficient G_{CCF} depicted in Figure 8.14(a) provides an indication about the richness of the overall VANET connectivity by averaging the local component coefficients ($L_{CCF}(i)$) of each individual vehicle. Specifically, it expresses the probability that two vehicles connected to a common vehicle, are also inter-connected themselves. On first sight, the VANET looks to be well-connected even in the early morning hours, where G_{CCF} has its lowest value of 0.67. Taking this instance as an example, if a vehicle i received a message from a neighboring vehicle j, then there is 67% probability that other vehicles within the communication range of i, will receive the same message from j.

Eventually, as the day progresses towards 7:00a.m and vehicular traffic around the city of Cologne becomes more dense, then the $GCCF$ gradually increases and stabilizes to 0.73.



(a) (b) (c)

(d) (e) (f)

Figure 8.15: Spatial representation of vehicle location correlated with: (Top) $LCCF$, (Bottom) $LCCF$ weighted with Degree. From left to right: $\tau = 1000s$, $\tau = 2000s$, $\tau = 3000s$.

Nevertheless, from the previous discussion regarding the size of components, we can see that 50% of the overall VANET components are comprised of 4 or less vehicles. Given that the spatial dispersion of vehicles participating in these components is small, the probability of cliques or almost-cliques being formed, is not negligible. Inherently, members of cliques are characterized by $LCCF_i \approx 1.0$, therefore one can claim that this high connectivity reported by the $GCCF$ behavior is biased. To investigate this, Figure 8.14(b) plots the $GCCF$, averaged over all degree-k nodes (x-axis in logarithmic scale). The diagram refers to the VANET at $\tau = 1500s$ and

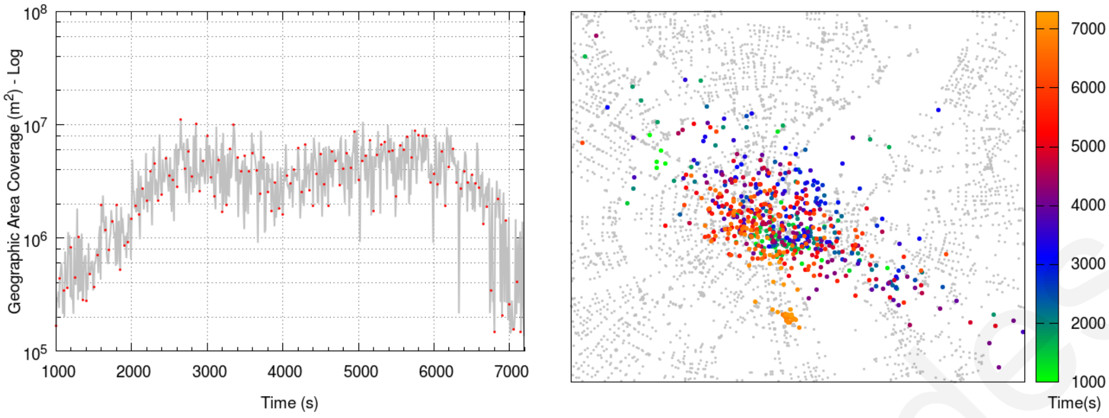
$\tau = 3500s$ with $GCCF$ equal to 0.63 and 0.69, respectively.[8] Evidently, vehicles with a low degree, which constitute the majority of the VANET participants, do not inflate $GCCF$. On the other hand, the small fraction of vehicles that exhibit high node degree are connected with other high degree vehicles, which themselves are also connected to other high degree vehicles. These findings can be visualized better through Figure 8.15. More specifically, the diagrams provide a spatial representation of vehicle location at different time instances. On the top row, vehicles are coloured based on their respective $LCCF_i$, alone. Given the predominance of red-to-orange coloring, overall it looks like that the vehicles tend to knit tightly among them, regardless of their position or the underlying road type. However, if the $LCCF_i$ is weighted with the degree of each vehicle, as depicted on the bottom row diagrams, we observe that strongly connected vehicles (and corresponding components) are found in the vicinity of road features that allow the formation of several wireless links and thus cause high degree. For instance, the various areas around the center where nodes are colored red, are major under/overpasses that are responsible to funnel traffic from freeways to arterials, or vice-versa. These ramps (often stacked in multiple levels) usually provide un-obstructed LOS to two or more high capacity road types and are often points of congestion. Subsequently, vehicles occupying them are presented with increased opportunities to establish inter-connections. The areas within the city center where red-color nodes exist are complex intersections with 3-ways or more.

Question #9: Where are giant clusters located? Does their location change in time?

Returning back to the study of the giant component, we would like to know where such structures can be found in the underlying topology and how their location shifts with time. Having such knowledge at hand is beneficial to communication protocols, mainly in situations where complete or partial VANET topology information is un-

---

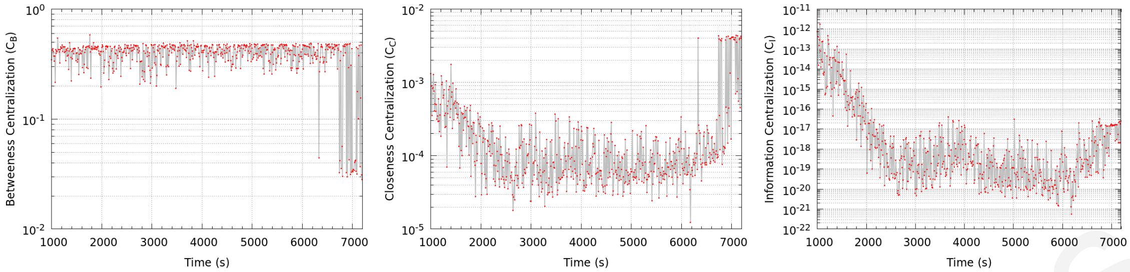
[8] To maintain the clarity of the figure, we present only the two particular time instances. However, this behavior is evident throughout the time window of our study.



(a) (b)

Figure 8.16: Temporal Evolution of Giant Component in terms of: (a) Geographic Area Coverage and (b) Coordinate Centroid.

available. Such details ultimately enable various protocols to approximate the rough geographic boundaries of the giant component, by employing only historical traffic statistics. Since the giant component is strongly connected internally, these approximations allow protocols to decide whether or not they are located within a region that favors multi-hop data dissemination. Figure 8.16(b) provides a visualization of the underlying road topology, overlayed by different points that correspond to the geographic centroid of the giant component. Here, the centroid metric is utilized since it can serve as an anchor point from which the giant component materializes. Note that each centroid point is colored in respect to time. During the early phase of the morning rush hour, the giant component evolves adjacent to arterial and collector roads that transfer commuters from the outer parts of the city and the suburbs towards the center. In these early times, the giant component centroids are spatially dispersed, and circulate both the eastern and western parts of the city where such road features can be found. The alterations between the different parts of the city, come as a result of the border effects in terms of component size, where at one particular instance a component at the eastern part of Cologne is larger by a small margin than another component at the western part, or vice-versa. When vehicle density is increased and



(a) (b) (c)

Figure 8.17: Centralization of the Giant Component. (a) Betweeness, (b) Closeness, (c) Information. As values tend to 0, the VANET becomes more homogeneous in terms of node centrality.

subsequently stabilized in the down-town area, the giant component centroids move in and remain concentrated in the center of the road network.

Question #10: What is the geographic area coverage of the giant component? How does it evolve in time?

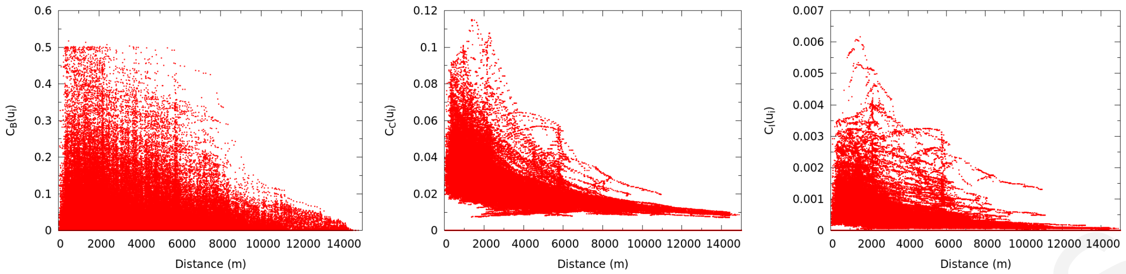
Figure 8.16(a) illustrates the geographic area coverage of the giant component across time. The coverage area is reported in m^2 and the y-axis is plotted in logarithmic scale. Given that the whole road topology size is $33Km \times 35Km$, the giant component can cover between 1% and 4% of the city of Cologne. The spatial coverage grows in the early phase of the morning rush-hour when vehicle density is increased. As in the case of the component membership, after the mark of ~ 7500 vehicles ($\tau = 1900s$) in the network is achieved, the geographic area coverage starts to stabilize with smaller variations. In contrast to smaller components in its vicinity that continue after this instance to change both size and coverage, the giant component remains relatively constant and always encapsulates instances of the four road features.

Question #11: How central are VANET nodes, and where central nodes can be encountered in the road network?

Considering the prominent presence and geographic coverage of this giant component, we then examine the importance and particularly the centrality of its member vehicles in the communication graph. Figure 8.17 depicts (y-axis in logarithmic scale) the centralization of the giant component for the three centrality metrics we consider in our analysis; betweenness (C_B), closeness (C_C) and information centrality (C_I), respectively. The centralization metric is a group-level metric that provides insights on the heterogeneity of the network. When the centralization attains its minimum value of 0, all members have exactly the same centrality index, that is the network is homogeneous. Conversely, as the centralization value tends to 1 the network becomes less homogeneous and thus more centralized.

With the exception of C_B, the remaining centrality measures exhibit a similar downwarding slope that conveys the temporal tendency of the giant component towards decentralization. Members of the giant component, are significantly heterogeneous in terms of their betweenness indices, especially prior the mark of $\tau = 2200$. During this early phase of the morning rush hour, we identify a number of vehicles who control up to 52% of the geodesics - that is, the shortest path(s) between nonadjacent vehicles - present in the giant component. As it can be seen from the respective plots in Figure 8.18, interestingly all of these high betweenness centrality vehicles are located on urban and collector roads in a distance of $\leq 6Km$ from the city center. Nevertheless, even during this period, vehicle mobility dynamics cause the giant component to alternate between states of centrality homogeneity and heterogeneity in a matter of a few seconds. As a result, vehicles that were previously identified as "central" to information communication within the network can often get demoted to ordinary nodes within 10-20 seconds, or vice-versa.

After $\tau = 1900s$, network centralization starts to decrease as a consequence of the giant component enlargment. Particularly, as its membership increases while being fused with smaller components in the vicinity, a number of new paths between the nonadjacent members of the giant component are introduced. Additionally, this restructuring process invokes the re-wiring of existing paths. Inevitably, the geodesic(s)



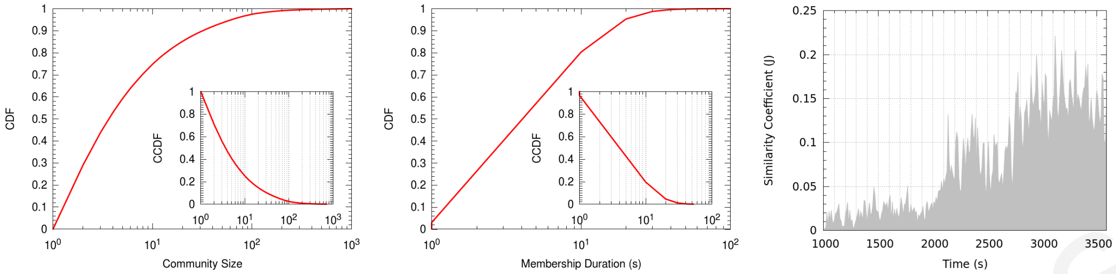
(a) (b) (c)

Figure 8.18: Centrality correlation with vehicle distance from city center: a) Betweeness, (b) Closeness, (c) Information. Data points are an accumulation of the 3600s we study.

between these nonadjacent vehicles have to go through additional intermediary vehicles, which themselves are also members of the giant component. Due to the fact that more and more vehicles now have some control over the information flows in the network, the betweenness score of these vehicles increases, in turn causing the global network centralization to decrease.

The giant component is quite homogenious in terms of closeness and information centrality of its members, since both of these metrics continuously exhibit a very low centralization ($\ll 0.1$). From the closeness perspective, there is a general trend that shows the inability of the network to support fast information spread, due to the large communication distances (hops and underlying geography) between vehicles connected via multi-hop links. Figure 8.18, shows the existence of a small number of vehicles that are somehow central since they have short distances to their nonadjacent neighbors. However, this effect decays almost exponentially as mobility takes them farther away ($\geq 2Km$) from the city center. Likewise, from the information centrality perspective, very few vehicles can be deemed as central when it comes to controlling information flows over all communication paths (including geodesics) and again they are restricted within an even smaller distance from the city center.

Question #12: Does the VANET contain communities?



(a) (b) (c)

Figure 8.19: Cumulative and Complementary Distribution Functions of (a) Community size and (b) Duration of continuous connection of "good" communities with the core. (c) Similarity coefficient of largest community.

We use the Girvan-Newman community detection algorithm [38]. We observe that the VANET exhibits strong modularity $Q_t \geq 0.92$ across time, indicating therefore a strong division of the network into communities with more dense inter-connections among their member vehicles. Specifically, 90% of the time instances we study indicate the presence of more than 560 distinct communities in the VANET. Nevertheless, as illustrated by the CDF in Figure 8.19(a), the majority of those communities are relatively small in size and rarely exceed 35 vehicles. According to the inset CCDF however, roughly 2% of the total communities detected in the VANET have more than 100 member vehicles and even so we note the existence of a single community that can grow significantly larger in respect to its counterparts and can reach size scales of up to 1400 members. A closer inspection using spatial analysis, indicates that this large community on most occasions is anchored near the city center where the majority of traffic is located. However, we noted several instances, where the location of this community shifts abruptly in a manner that minimally covers the area around the center.

Interestingly, the VANET satisfies the community profile of real-world networks, such as the Power Grid, that are embedded into low-dimensional geometric structures (i.e surface of the earth). According to such networks that were studied extensively in [80], "good" communities in the VANET - that is, highly inter-connected sets of vehicles - have a size scale of roughly 200 nodes, with this finding remaining quite consistent throughout time. These smaller communities are tenuously connected to the core of the VANET, oftentimes with just 1-2 wireless links. As the CDF in Figure 8.19(b) illustrates, the duration of a continuous connection of such peripheral communities with the core, rarely exceeds 20 seconds. Nevertheless, their dense structure allows them to act (even in isolation) as "data-islands", facilitating thus spatio-temporal information replication and maintenance during broadcast of geographic hovering communications.

As communities cross over the size scale of 200 nodes and gradually get larger and larger, they become less and less community-like (i.e. transition to sparsely connected state) and fuse with the network core. However, this transition to a sparsely inter-connected state, does not negate the fact that such communities can indeed continue supporting communication and be effective spatial information maintainers. Figure 8.19(c) depicts how similar[9] in terms of vehicle membership, is the largest community across time. We observe that after a particular traffic density is reached, more than 5% of its member-vehicles are constantly the same in between consecutive time instances. Hence, once identified, these are the vehicles that can be entrusted with the task of information storage and replication for a particular time frame.

Question #13: What are the lifetime characteristics of links in the VANET?

Link level analysis of the VANET communication graph contributes to the prediction of the network-link lifetime. The number and duration of connected periods between any two vehicles, as well as the duration between successive connected periods is influenced by microscopic mobility characteristics as well as the underlying road network. To this end, Table 8.3 present the link-level statistics of the VANET members in the city of Cologne.

---

[9] Similarity here is obtained by iteratively calculating the Jaccard similarity coefficient of the largest communities between two consecutive time instances.

Cologne VANET				
	Min	**Max**	**Mean**	**Median**
Number of Connected Periods	1	8	2.2	2
Link Duration	1 sec	560 sec	322.7 sec	301 sec
Re-Healing Period	1 sec	980 sec	121.16 sec	117 sec

Table 8.3: Link level statistics.

Intuitively, higher traffic densities result to smaller spatial inter-vehicle distances and thereby increase the time period in which two vehicles are in range of each other. Consequently, this allows established wireless links to have a longer duration. An increased link duration reduces the number of connecting periods between the two vehicles and consequently minimizes any overhead imposed by the process of re-establishing connections. Higher link durations are therefore encountered particularly in arterial and freeways, where the characteristics of these roadways do not cause links to break often. We observe that the number of connected periods for vehicles being driven on urban and collector roadways is almost 3x as their above counterparts, however with smaller re-healing durations.

Chapter 9

Urban Traffic Sensing and Acquisition

Here we lay the groundwork for *V-Radar*, a vehicular traffic information query protocol for urban environments based on V2V communication. In contrast to other approaches in the literature that aim to obtain traffic information on *singular roads*, V-Radar enables the querying and acquisition of traffic information along a composite *road-path*, starting from a vehicle's current position towards its final destination. Specifically, V-Radar is able to query not only the initially selected road-path, but also a number of alternate paths that lead to the vehicle's destination. This allows the driver or the in-vehicle navigation system to establish a more broad and complete view of the traffic conditions that will be encountered further ahead. Such knowledge can be extremely valuable in the process of calculating a more optimal route to the destination in terms of travel time. Furthermore, this work proposes a number of components for V-Radar that can work in tandem so as to maximize the number or road-paths monitored, whilst keeping wireless transmissions and bandwidth utilization at the minimum. Specifically the contributions of this work are:

- We define the problem of identifying in real-time, in the absense of an infrastructure navigation service, the set of road-paths among any two road intersections, that if followed will result in reduced travel times. To identify such road-paths we utilize location-dependant queries.
- We propose V-Radar, a traffic information query protocol for urban environments using V2V communications. The advantage of V-Radar over related works is its ability to monitor the prevailing traffic conditions in a number of road-paths from a vehicle's current location towards its final destination using location-dependant queries.
- Using realistic trace-driven simulation studies, we show how V-Radar even in its simplest form has a significant performance advantage over existing vehicular traffic query methods available in the related literature.

Information regarding the underlying road-network topology is provided to the vehicle through on-board preloaded digital maps, while its current geographic location is obtained via the Global Positioning System (GPS). Such maps are enriched with historical statistics that exhibit the traffic conditions of the road-network at different times of the day.

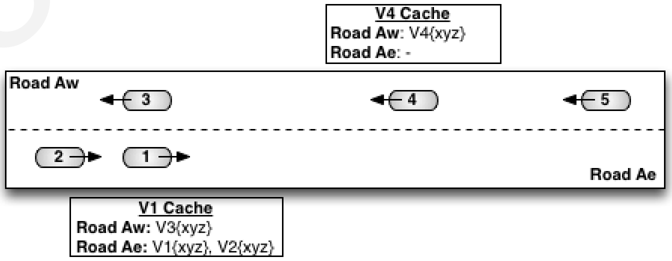


Figure 9.1: Caching Sensed Traffic Information

Each vehicle encompasses a local cache where it temporarily maintains vehicular traffic information sensed through received beacons from the roads that have been traversed. For example, as shown on Figure 9.1, two new records are created in each vehicle's cache upon entering a new road. One record is for Road segment A_e and one for Road segment A_w.[1] Whenever a beacon is received from a neighboring vehicle, the respective road records are updated based on the value of rid - vehicle's 1 cache contains entries for both Road A_e and A_w through the beacons of vehicles 3 and 2 respectively.

Obviously, the decision to follow the shortest road-path(s) in terms of geographic distance does not necessarily guarantee a reduction in the total commute time towards a specific destination, since the prevailing traffic conditions might dictate otherwise. For instance, if several roads in the shortest geographic road-path exhibit high vehicle density and low average speed, this will eventually result in a higher total travel time than if following another road-path with sparse traffic.

Therefore, an IVC-enabled vehicle can provide its driver or the on-board navigation system with the necessary traffic information required to make the aforementioned decisions by issuing location-dependent traffic queries to other vehicles.

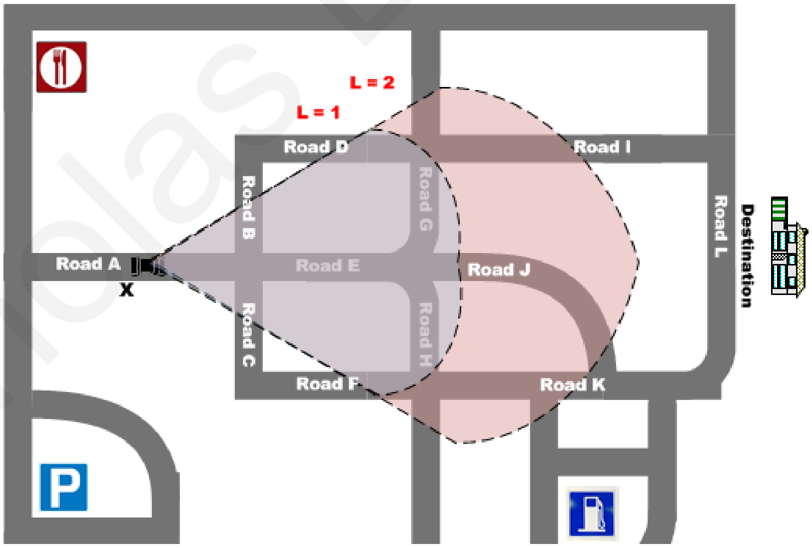


Figure 9.2: Querying for Traffic Information using V-Radar

---

[1] The subscript next to each road denotes traffic direction: "e" for eastbound, "w" for westbound.

9.1 V-Radar: A Vehicular Traffic Query Protocol for Urban Environments

The aim of V-Radar protocol is to support the generation, dissemination and resolution of such location-dependent traffic queries in a manner that resembles a forward-scan radar, similar in concept to the radars used in maritime or aviation activities to detect the presence of obstacles that might block the normal course of a vessel or an aircraft. In the context of V-Radar such obstacles are considered to be traffic build-ups or congestion, which as aforementioned can cause severe delays.

The concept of V-Radar is explained with the following example. We assume that vehicle X on Figure 9.2 is driven eastbound on *Road A* and wishes to follow the road-path to its destination Y with the minimum travel time. We also assume that X currently follows the shortest geographic path towards Y suggested by the on-board navigation system. However, X would also like to know the prevailing traffic conditions for the other available road-paths (with larger travel distance) to Y, in case a route change is required. Through its knowledge of the road network, the on-board computer of X can calculate and consequently rank the K possible road-paths in terms of the geographic distance from the next intersection to the intersection which is closest to Y.

Upon entering *Road A*, X issues *LookAhead (L)* traffic information queries at a selected rate (R) towards all the calculated K road-paths. We envision *LookAhead* queries as radar pulses that travel towards a specific direction and as soon as they hit on a surface (i.e. a moving or stationary object) they are reflected back towards the source. *Hence, LookAhead queries are propagated to a certain depth L in each of the identified road-paths and collect the traffic conditions of all the roads up to and including the specified depth.*

Therefore, if X would like to know the traffic conditions in all the road-paths up to 2 roads ahead of its current position, then the look ahead value will be set to $L = 2$ and the following traffic queries (Q) will be generated: $Q_1 : \{Roads\ B, D, I\}$, $Q_2 : \{Roads\ E, G, I\}$, $Q_3 : \{Roads\ E, J, K\}$, $Q_4 : \{Roads\ E, H, K\}$ and $Q_5 : \{Roads\ C, E, K\}$.[2]

Consequently, each query is propagated in a multi-hop fashion to each individual road in a given path, where the required traffic information is retrieved either with on-the-fly cooperation of other vehicles on location (i.e. the concept of VAHS in [42]) or from vehicles' cache. Upon reaching depth L and retrieving the required traffic information query replies are routed back to X. This query cycle process iterates until X reaches its destination Y.

9.2 Problem Formulation

A road-network can be considered as a directed graph $G = (V, A)$, where intersections or end-points (dead-ends) correspond to the set of vertices's $V = \{v_i\}$ and roads to the set of arcs $A = \{a_{ij}\}$. Two nodes v_i and v_j can communicate directly with each other, if they are connected by an arc a_{ij} and no other intersection or end-point exists in-between them. A path P_{ij} is a unique, alternating sequence of connected nodes and arcs in G that starts from v_i and ends at v_j. $D(P_{ij})$ is the length of path P_{ij}. Consequently, P denotes a set of paths that can be defined in $G = (V, A)$ from v_i to v_j. To this end, a road-path is considered as the list of roads - and by definition, intersections - that a vehicle has to traverse in order to travel from one intersection to another intersection. Taking under consideration the historical traffic statistics, each arc is attributed with the average speed $\bar{u}$ and average vehicle density $\bar{\rho}$ of the respective road.

---

[2] Note: Although overlapping road-paths are possible and will be identified (i.e $\{B, D, I\}$ and $\{B, D, G\}$), in the example we deal only with distinct, non-overlapping paths.

Let $t_{P_{ij}}$ be the time it takes to travel from node v_i to node v_j in any path P_{ij}. Since travel times are influenced from static parameters such as the road length but also by the dynamic traffic conditions (average speed, vehicle density, traffic light queues, etc.) vehicles would like to identify the set of road-paths from node v_i to node v_j $S_{ij} = \{P_{ij}^1, P_{ij}^2, P_{ij}^3, \ldots, P_{ij}^K\} \subseteq P$ towards their destination such that: $t_{P_{ij}^n} \leq t_{P_{ij}^{n+1}}$, for any $n \in \{1, \ldots, K-1\}$.

A naive assumption would be that to discover these road-paths and construct S_{ij}, a vehicle should firstly calculate *all* the possible geographic shortest paths and subsequently generate and transmit the necessary traffic queries towards them so as to acquire the necessary information for the calculation of $t_{P_{ij}}$. However, due to various constraints such as query-reply delay thresholds, size of the road topology, vehicle density and the VANET connectivity status, querying the traffic conditions of: i) all paths $P_{ij} \in P$ and ii) consequently all roads (whole length D) of any individual path P_{ij}, is by no means scalable and with very high probability will lead to inaccurate or out-of-time results. Therefore vehicles should be able to estimate the number K of paths to be queried and additionally up to what depth L of each P_{ij}, *LookAhead* queries should be propagated in order to meet constraints such as the above.

Moreover, due to the high dynamics of vehicular environment where the conditions on any given path can change abruptly, one may assume that a high query generation rate R would be required to capture traffic accurately. However, in an urban setting where a large number of vehicles are continuously competing for the VANET resources, it is crucial to be prudent in the use of the wireless channel and thereby refraining from transmissions in a selfish manner.

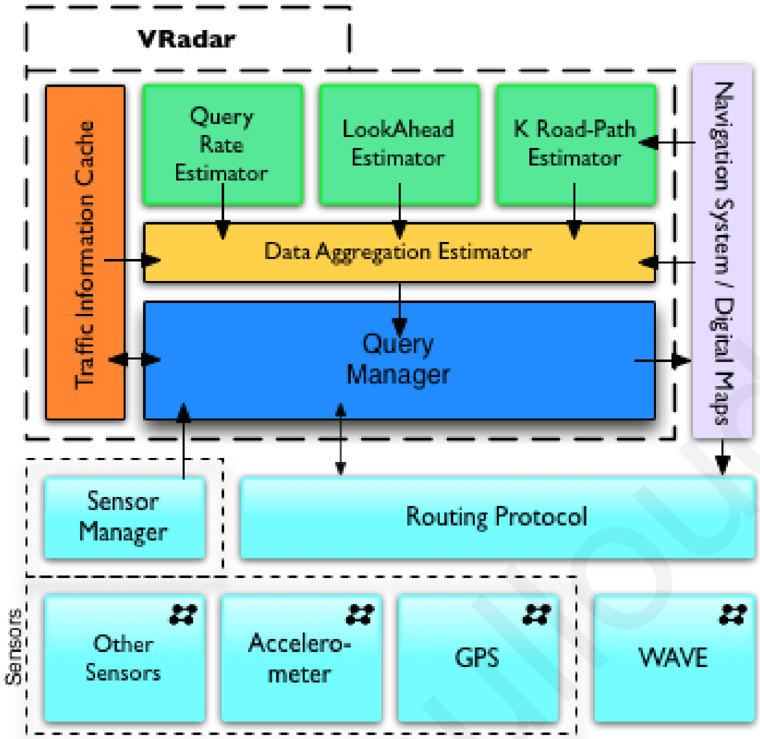


Figure 9.3: V-Radar Architecture

9.3 Objective

The objective of the V-Radar query protocol is thus to provide the necessary mechanisms that enable: *i) the sustainability of an acceptable traffic query-reply delivery rate and ensuring any delay thresholds are met, ii) the maximization of the number of alternate road-paths* ($max(K)$) *and roads* ($max(L)$) *to be monitored, whilst iii) minimizing wireless transmissions and bandwidth utilization.*

9.4 V-Radar Architecture

This section introduces the V-Radar architecture. Its modular design is such that it allows the use of various components that can collectively realize the V-Radar objective. As Figure 9.3 depicts, V-Radar runs on the application layer of the vehicle's on-board computer therefore it can utilize the underlying routing and transport protocols. Below the function of its individual modules is explained:

Road-Path Span (K) Estimator Module

It is evident that monitoring the traffic conditions of all available paths towards the destination does not scale in the real-world. Besides the unknown number of road-paths which might be available, the resulting overhead that will be imposed in the VANET from such an attempt will be forbidding. The Road-Path estimator function is to identify the critical value K, that is which and how many of the available road-paths towards the destination will be queried. It interfaces with the Navigation system in order to be aware of the road topology and various driver preferences (e.g. maximal deviation in terms of geographic distance from the initially selected road-path, willingness to use road-paths with bridges or tolls, etc). The K estimator can additionally utilize information items from the Traffic Information Cache (see below) to assist in the estimation process. For instance, knowing from past queries that the traffic conditions of a particular road-path exhibit variability below a certain threshold, then that road-path can be queried on a less frequent basis.

LookAhead (L) Estimator Module

The L value and specifically how deep a road-path will be queried has a crucial role on the correctness of the information that V-Radar will provide to the driver. Querying too shallow might lead to a horizon effect where a congestion further down the path might not be identified. On the other hand, querying too deep in the path introduces the risk of a packet loss due to network fragmentation or the violation of query-reply delay thresholds.

Query Rate (R) Estimator Module

To maintain an update view of the traffic conditions in the identified K road-paths, new query messages need to be dispatched periodically. Nevertheless, the new query generation rate depends on the existence of several constraints that must be taken under consideration. For instance, vehicle speed and/or current road length impose a time constraint on when the next junction will be reached. In a higher vehicle speed and small road length setting, queries need to be generated at an increased rate. On the other hand, consideration needs to be given on the number of neighboring vehicles that compete for the wireless medium and which can influence the number of message lost due to possible packet collisions.

Data Aggregation Module

It is possible that several of the identified K road-paths will be overlapping. In such cases "duplicated" query messages will be generated, causing unnecessary utilization of the wireless medium. The function of the Data Aggregation module is to identify such situations and provide the necessary mechanisms that facilitate the aggregation of query information in a single message such that duplicates are avoided.

Traffic Information Cache (TIC)

The cache module is used to store traffic information (average speed and vehicle density) about various roads. Such information is obtained as explained in the System Model or through the contents of a received query-reply. Each cached item is a 4-value tuple *[id, timestamp, data-type, value]*. Specifically, the index value $id = rid$, *timestamp* determines the point in time where the provided information becomes stale, *datatype* is the type of cached information (speed, density, etc) and *value* is the actual value of the sensed or received information. A Cache Replacement Module (CRM) is used to define and enforce the necessary policies (e.g. LRU, MRU, etc) that dictate the eviction of stale items and maintaining the cache freshness.

Query Manager

The Query manager orchestrates the operation of the V-Radar protocol components on each vehicle. Specifically, it collects the values calculated by the K, L, R estimators and is responsible to construct the respective V-Radar query messages for each of the identified road-paths. Furthermore, it is responsible for resolving incoming traffic queries.

[Bytes]

1	2	3	4
Type	C ; R ; A ; ?	*LookAhead*	*MsgSize*
Message Sequence UID			
Lifetime			
Originator Address			
Road[0] UID			
Traffic Information for Road[1]			
. . .			
Road[L-1] UID			
Traffic Information for Road[L-1]			

Table 9.1: V-Radar Query Message Format

As Table 9.1 illustrates, the V-Radar query message consists of a 16-byte fixed header and a variable-sized payload (data) section. The header fields have the following functions:

- *Type:* Indicates what type of traffic information the query message contains. For instance, a value of 1 can be used for average vehicle speed, while 2 for vehicle density.

- *Boolean Bit Flags:* The "C" flag indicates whether cached traffic data can be appended to the reply message. The "R" flag indicates whether the message is a query-reply. A query-reply message is considered to be the message which includes traffic information for at least one road in the road-path being queried. The "A" flag indicates whether a receiving vehicle is allowed to append traffic information to the message. The remaining bits are reserved for future use.

- *LookAhead:* Indicates at what depth in the road-path being queried the message currently is.

- *MsgSize:* A count of the total number of bytes contained in header and data sections. As the header length is a fixed size, this field effectively tracks the length of the variable-sized payload. In effect it indirectly reports the number of roads in a path that the message will visit.

- *Sequence ID:* A sequence number uniquely identifying the particular query message when taken in conjunction with the originating node's address.
- *Lifetime:* The time in milliseconds for which vehicles receiving the query message consider it to be valid and are allowed to forward it or append traffic information to its payload section.
- *Originator address:* The address of the node from which the query message originates.

The payload section of the V-Radar query message contains (L), 16-byte fields. The first 4-bytes of each pair are used for storing *rid* of the roads in a path to be queried. The sequence of *rid* in the message is also indicative of the order that each road in the path will be traversed by the query. The remaining 4-bytes are used for storing the queried traffic information for the respective road.

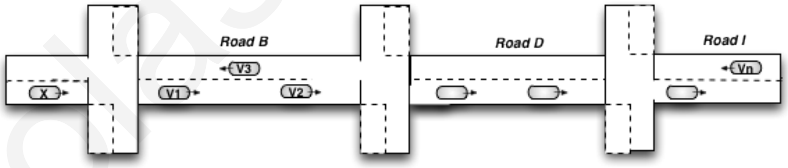


Figure 9.4: Query Resolution Illustration

9.5 Query Resolution

Figure 9.4 depicts how a V-Radar query is generated and resolved in the context of the example presented earlier in the system model. Vehicle X selects to query and obtain the average vehicle speed on road-path $P_1 : \{Roads\ B_e, D_e, I_e\}$. V constructs a new query message Q (as per Table 9.1), inserts in the payload the UID's of the roads to be queried and sets flags $C = 1$ and $A = 1$. Once the V-Radar Query manager in V_1 receives Q, it parses the message and consults the Navigation System whether the vehicle's current position qualifies participation in the query resolution. Since V_1 is the first vehicle on $RoadB_e$ that can join the query resolution, the manager updates the LookAhead field in the header to reflect the current road-path depth ($L = 1$). Consequently, it piggybacks its current speed in the appropriate payload field for $RoadB_e$ and passes the message on to any available neighbor. If Q is received by a vehicle on $RoadB_w$ (V_3), the query manager aggregates traffic information for $RoadB_e$ stored in the TIC and appends them to the query. It then sets flag $A = 0$ to avoid duplicate information from other vehicles further down the road (e.g. V_2). Once Q approaches the end of $RoadB_e$, the receiving vehicle checks if it is allowed to append its own speed information, sets flag $R = 1$ to indicate that the query contains traffic information for at least one road, sets flag $A = 1$ and forwards the message to $RoadD_e$ where the above process is repeated. In the case that Q is received by a vehicle such as V_n, where TIC information indicate that no vehicle was sensed in the opposite traffic direction (i.e average speed for $RoadI_e = 0$), the query manager may decide to forward Q back to V based on various rules. Finally, when Q traverses the last road in P_1 and records the necessary information, it is routed back to vehicle X.

9.6 V-Radar Evaluation

9.6.1 Simulation Setup

This section presents how V-Radar in its simplest form performs against other related works. For the purposes of the evaluation, V-Radar was implemented as an application module under the ns-3 [23] network simulation framework. Since we are still researching on the techniques that will eventually be used for the estimators modules presented in Section 9.4, in the following evaluation we are simulating their existence

Vehicle Transmission Range	300m (802.11p)
Simulation Time	1000s (200s warm-up)
Routing Protocol	VADD (H-VADD)
Beacon Generation Rate	10Hz with some jitter
Query Generation Rate (pkt/sec)	0.1 - 1.0
Query TTL	64 hops
Cache Items TTL	128s
Query Sources	200
K-Paths	3

Table 9.2: Network Simulation setup parameters

by using different values for K, L and R. We compare V-Radar against the road network based adaptive query evaluation (RNBAQ) method proposed in [53].

Realistic urban mobility traces were utilized from an improved version [115] of the TAPAS-Cologne [24] dataset. We extract the mobility traces from the TAPAS-Cologne dataset for all vehicles moving within a 4km X 3km rectangular area surrounding the Cologne city center.

200 randomly selected vehicles issue queries for 800s. At each intersection each vehicle calculates the Top-K shortest paths towards its destination with $K = 3$. Each road-path is monitored up to a depth of 7. Therefore, depending on the experiment, queries can take a value $1 \leq L \leq 7$. For V-Radar, simple information caching is used. A vehicle can provide a reply to a received message if it knows traffic information for one or more roads that will be queried. Cached items TTL is set to 128sec. For routing queries to destination roads, we utilize VADD [126] as the underlying routing protocol. For RNBAQ, we extend VADD to facilitate the broadcast of control messages used for location change notification. Each vehicle broadcasts a HELLO beacon at a rate of 10Hz. Data-rate was set to 3Mbit/s, and all the PHY and MAC properties conform to IEEE 802.11p [63].

Note that the results depicted in this section, account for the average values of each of the studied metrics, calculated over 20 runs for each simulation scenario, with different random number seeds. Finally, Table 10.1 presents all the parameters that were utilized in the simulation.

9.6.2 Evaluation Results

The following metrics are used in the evaluation:

- *Packet Delivery Ratio:* the ratio of queries successfully delivered back to the source vehicle to those generated by the source vehicle.
- *Network Overhead:* the total number of KBytes transmitted. The total number of KBytes is inclusive of any control messages and the underlying routing protocol headers.
- *Accuracy:* denotes whether the retrieved traffic information is close to the real value. Real values are obtained from traffic statistics computed by SUMO.

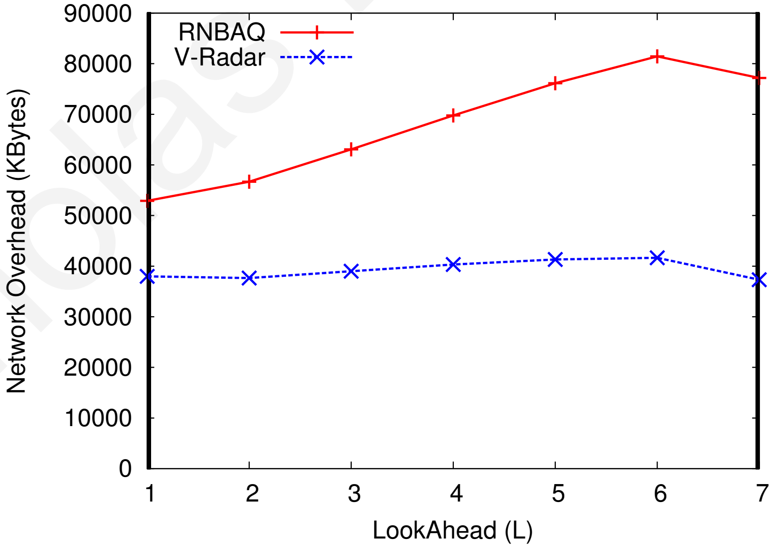


Figure 9.5: Network Overhead vs LookAhead

As depicted by Figure 9.5 the network overhead imposed on the VANET by the V-Radar query protocol is substantially lower than the in the case of RNBAQ. While

V-Radar uses a single packet for querying the traffic conditions of all the roads in a path, RNBAQ is required to generate and inject in the network one packet per road. In addition, due to its single packet/query design, V-Radar keeps the network overhead lower than RNBAQ as the query generation rate increases (Figure 9.6).

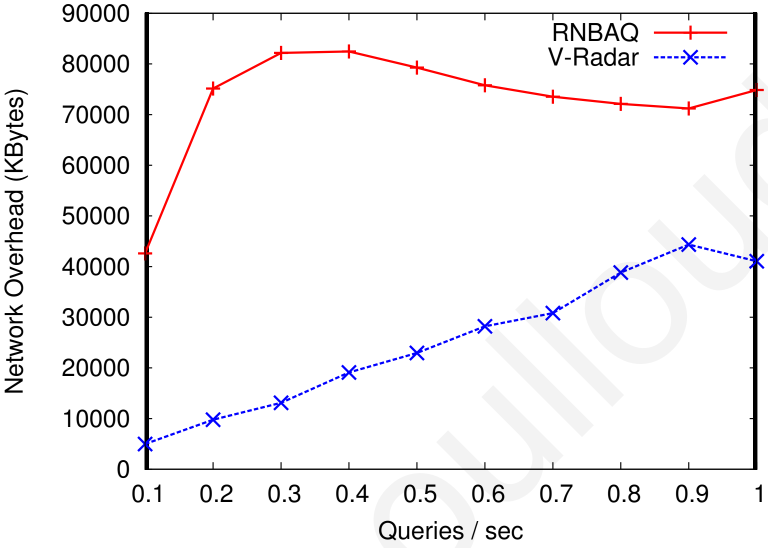


Figure 9.6: Network Overhead vs Query Generation Rate

Nevertheless, the query generation rate depends on various constraints such as query-replay delay thresholds and minimum information accuracy levels. It is crucial, therefore to investigate intelligent techniques that can be utilized the V-Radar for estimating the query rate R in order to provide even better utilization of the wireless medium.

By examining Figure 9.7 we can observe that as the number of generated queries increase (due to the LookAhead parameter increase), the PDR of RNBAQ drops faster than in the case of V-Radar. This behavior can be attributed to packet collisions that take place at the busy wireless channel and cause several traffic query messages to be lost. Unavoidably, V-Radar also suffers from packet collisions. However, the inherently smaller number of generated messages mitigates the above side-effects.

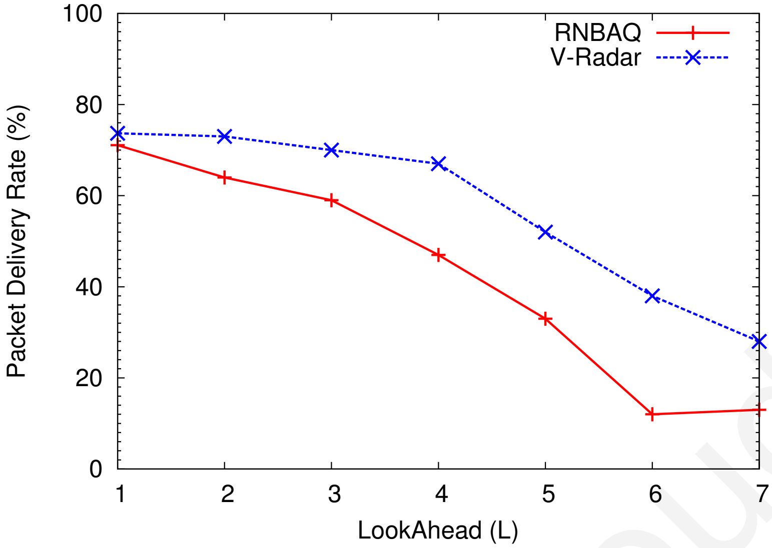


Figure 9.7: Packet Delivery Rate vs LookAhead

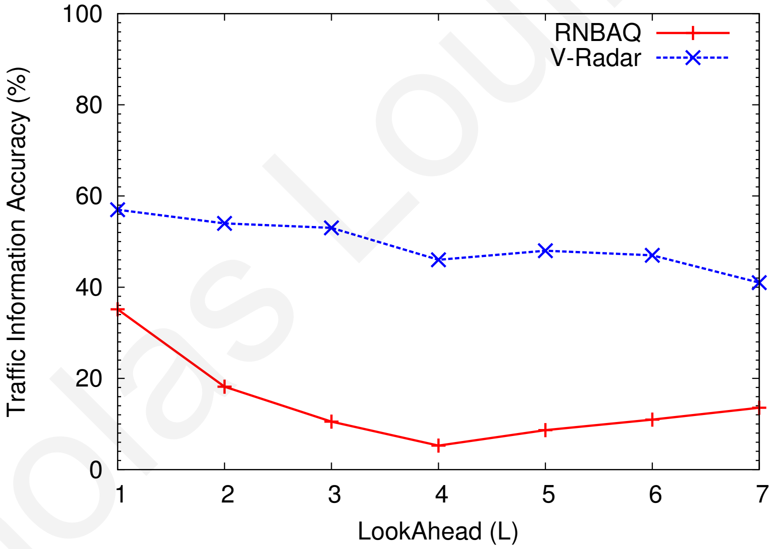


Figure 9.8: Traffic Information Accuracy vs LookAhead

Moreover, the improved PDR that V-Radar exhibits over RNBAQ is attributed to the existence of the Traffic Information Cache. There is a high probability for a query message to encounter an intermediary vehicle on a monitored road-path that can provide traffic information for one or more roads from its local cache. In turn, this overcomes the need for a query to traverse the whole road-path up to L in order to retrieve the required traffic information. Therefore, it enables the query return back to the source vehicle prior to the message TTL expiration.

Although the retrieved traffic information accuracy is better than in the case of RNBAQ, Figure 9.8 indicates that it remains quite low and becomes lower as the LookAhead value increases. However this is expected since it takes more time to reach a road further away and return the result back to the source vehicle. Subsequently, the returned results do no reflect correctly the real situation in the road network. Here, clever traffic information caching techniques can be utilized in the vehicle cache which will allow a query to be answered faster and with better accuracy. Therefore it is a clear indication that the design and implementation of an adaptive CRM for V-Radar is important.

Chapter 10

Large-Scale Traffic Information Diffusion using Reactive Mechanisms

10.1 Performance Evaluation of VANET Routing Protocols in Large-Scale Urban Environments

In this work, we evaluate and compare the performance of three highly established VANET routing protocols GPCR [82] (multi-hop protocol), VADD [126] (carry-and-forward protocol) and LOUVRE [76] (overlay protocol), under different urban scenarios of varying size and realism. These protocols are cited in many research studies and are known as good performers in their respective classes of routing protocols [32, 37, 81]. Initially, we examine their performance following simplistic approaches undertaken in the majority of the literature so far. Consequently, we combine highly realistic vehicle mobility in a large-scale urban topology as well as network

	GPCR	**VADD**	**LOUVRE**	**AODV**
Protocol Variation	GPCR-CC	H-VADD	N/A	N/A
Vehicle Transmission Range	300m (802.11p)			
Propagation Model	Nakagami Propagation Loss			
Simulation Time	1000s (200s warm-up)			
Data Rate	3Mbit/s			
Beacon Generation Rate	10Hz with some jitter			
Packet Generation Rate (pkt/sec)	0.2, 0.4, 0.6, 0.8, 1.0			
Packet Size	1KB			
Packet TTL	64			
Cache TTL	128 sec (where applicable)			

Table 10.1: Routing Protocol Evaluation - ns-3 Simulation Parameters

traffic generated from an exemplary traffic query application. By doing so, we aim to evaluate their performance under conditions that strive to resemble as closely as possible the behavior and the environment that each single car would face in reality. We argue that results stemming from such a realistic and complete scenario increase the possibility of identifying problems as well as implications in the design of routing protocols that need to be considered and addressed for achieving optimal performance. To the best of our knowledge, this is the first research work in the related literature that not only brings under one roof three highly established routing protocols that have been proposed specifically for VANETs, but also evaluates their performance in a realistic manner.

10.1.1 Performance Analysis

10.1.2 Simulation Setup

For the purposes of the evaluation, GPCR, VADD and LOUVRE were implemented from scratch under ns-3.11 [23], trying to remain as accurate as possible given the information provided in the original articles [76, 82, 126]. For baseline comparison, we include in our evaluation the well-known AODV protocol [101]. AODV is a MANET reactive routing protocol that builds routes on demand, and is often used in the literature when evaluating the performance of VANET routing protocols. The reasoning behind using AODV here, is to depict to the reader how the performance of a MANET protocol compares against the performance of VANET dedicated protocols in the vehicular environment. Based on the findings of [40], each vehicle broadcasts a HELLO beacon at a rate of 10Hz. Data-rate was set to 3Mbit/s, and all the PHY and MAC properties conform to IEEE 802.11p [63]. Table 10.1 presents all the parameters that were utilized in the network simulator throughout the evaluation.

10.1.3 Examined Metrics

We reside in the following metrics in order to evaluate the performance of VADD, GPCR and LOUVRE.

- *Packet Delivery Ratio:* the ratio of queries received by the destination vehicles/sites to those generated by the source vehicles.
- *Number of Hops:* the average number of vehicles a query has traversed in order to reach the destination. For round-trip queries, the total number of hops (source-destination-source) are calculated
- *Average Delay:* the average difference between the time a traffic query was generated by the source node and the time the reply to the source node was received. Dropped or lost queries are not included.

Note that the results depicted in this section, account for the average values of each of the above metrics, calculated over 5 runs for each simulation scenario, with different random number seeds.

We employ the following scenarios to determine to what extend the road topology, vehicular mobility and the application (information exchange model) affect the performance of the above VANET protocols.

10.1.4 Typical Scenario

Initially, to understand to what extend the road topology and consequently vehicular mobility affect the performance of the above protocols, we opted to experiment on a scenario typical to what is used in a number of research works in the bibliography. The scenario consists of a 4x3Km grid-layout road topology extracted from the U.S. Census Bureau TIGER [27] database, with 18 intersections and 26 bi-directional roads. Intersections are not controlled by traffic lights, therefore vehicle turns (straight or left/right-turns) are dictated by the prevailing traffic conditions at the intersection. To simulate vehicular traffic conditions similar to that of an urban environment, all horizontal roads are set as high-speed roads with a speed limit of 80Km/h, while all vertical roads are set as local roads with a speed limit of 55Km/h. We feed the road network to SUMO [21] and generate random trips for 250 vehicles, making sure that all of them remain in the map for the whole of the simulation area. We provide two variations of this typical urban scenario, one using a simplistic information exchange model that mimics a parking place reservation application and one using the V-Radar application. Below we present the results of these two variations.

Parking Place Variation - we assume that a number of vehicles would like to make a reservation to a specific parking-lot. Four static nodes are placed at the corners of the grid road topology to simulate such a site. Among all vehicles, 15 of them are randomly selected to send Constant-Bit Rate (CBR) data packets to these 4 static sites. No reply is send back to the vehicle for a reservation and duplicated requests arriving at a site are simply discarded. We perform different simulations to study the effect of varying the data sending rate (as per Table 10.1).

It can be observed in Figure 10.1, VADD achieves the highest packet delivery ratio among all protocols. This is primarily due its capacity to cache packets (a process known as carry-and-forward) when: (i) no other vehicle exists in the vicinity of the current packet carrier (i.e. network fragmentation), or (ii) none of the carrier's existing neighbors is considered to be a better candidate to uptake the role of forwarding the packet to its destination. Specifically, in this occasion, where the underlying road topology layout and sparse vehicle density constantly drive the network to fragmentation, VADD performs better by buffering packets until network connectivity is established. On the contrary, GPCR and LOUVRE do not support carry-and-forward, hence they exhibit a much lower packet delivery ratio. When such of the aforementioned conditions are encountered in the network, both GPCR and LOUVRE silently drop the packet.

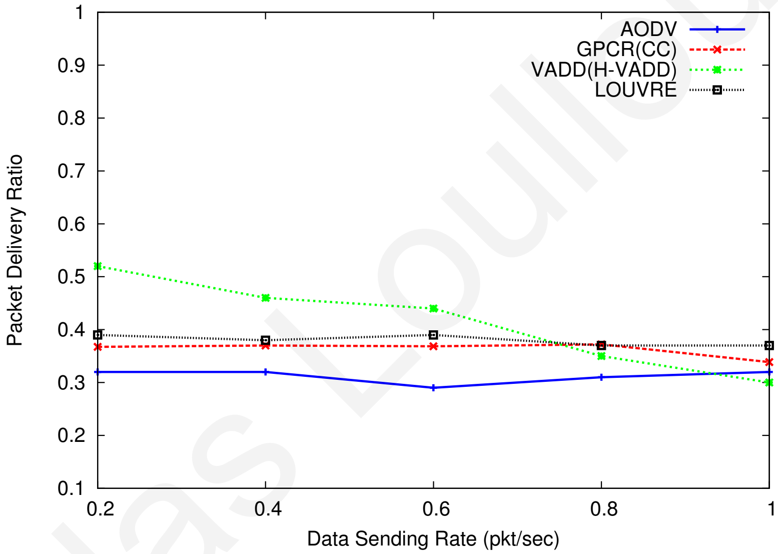


Figure 10.1: Packet delivery ratio as a function of packet sending rate - Typical scenario for parking place reservation

The substantially higher packet delivery delay that VADD exhibits in Figure 10.2, in contrast to the delay imposed by the other three protocols, is attributed to the additional time a packet spends in a vehicles cache. Our study revealed that the increased packet delay in conjunction with the better packet delivery ratio seen previously, stems from the fact that VADD can serve packets for which their origin vehicle is geographically further away from the destination site. Similarly, the very low packet delivery delay ($\leq$1 sec) exhibited by GPCR, LOUVRE and AODV, is due to the fact that all successfully received packets were the ones that either got generated at proximity of the destination or on the very few times where mobility was such that it permitted end-to-end connectivity between geographically distant nodes.

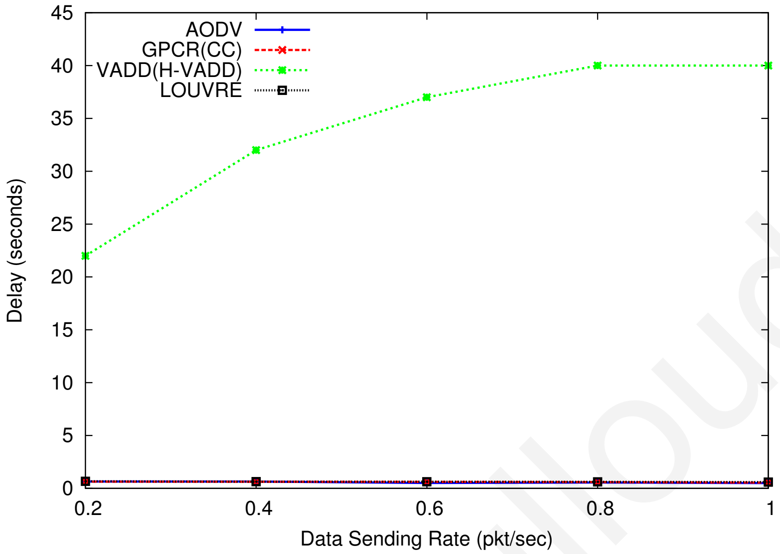


Figure 10.2: Packet delivery delay as a function of packet sending rate - Typical scenario for parking place reservation

As depicted in Figure 10.3, data packets in VADD traverse on average the same number of hops as packets that are routed with the aid of GPCR, LOUVRE and AODV. This is a clear indication that carry-and-forward mechanisms do not affect the number of hops during an end-to-end communication, they merely assist by enabling the bridging of network fragments. We note that the average 5-7 hops that a packet is required to traverse en-route from source to destination are indicative of the dimensions of the respective underlying road topology and a wireless transmission range of approximately 300m.

Implications: From the above findings, it is evident that in a scenario such as the parking spot reservation where there is delay tolerance, carry-and-forward protocols in the likes of VADD are highly suitable. Such protocols increase the probability that a request will be propagated to its destination even with some acceptable delay. Most importantly though, they indicate that the performance of non carry-and-forward protocols (i.e GPCR and LOUVRE) can be underestimated due to conditions of poor network connectivity induced either by low-vehicle density (considering the 250 vehicles in a 4x3 Km area) or unrealistic mobility. These two factors can cause network fragmentation, which inevitably will result in packet drops and consequently low packet delivery ratio.

V-Radar Variation - In light of the above we extend the evaluation of the typical urban scenario, by introducing a second case, where all vehicles are installed with the V-Radar traffic information application presented in Section 9.1. Prior to initializing the simulation, all vehicles are pre-loaded with a map of the underlying road topology (a directed graph) and traffic statistics computed by SUMO.

In the V-Radar scenario, once a vehicle enters a new road, it starts generating queries (with *LookAhead (L)*) to all roads that make up the available road-paths leading to its destination in order to identify the prevailing traffic conditions. A vehicle calculates all the K-shortest road-paths from its current position to its destination by running Yen's K-shortest path algorithm [122] on the road topology directed graph. For the purposes of this evaluation we select a value of $K = 3$ and $L = 2$. In addition, data packets containing traffic information queries are generated with a CBR rate(R) between 0.2 and 1.0 packets per second (incremented by 0.2).

Once a vehicle approaches the end of the road, it stops generating queries until it crosses an intersection and enters a new road. Queries are propagated to the destination roads, where a reply is generated and immediately routed back to the source node. In case that no vehicle exists in any of the destination roads, the packet is dropped. This road query process is performed continuously, until 100 seconds prior to the simulation end in order to allow for all vehicles to process any packets which are still propagating the network (queued queries or replies en-route to the source vehicle).

Implications: In contrast to the previous case where all vehicles were sending packets to a fixed site, here we observe the performance of all protocols to degrade

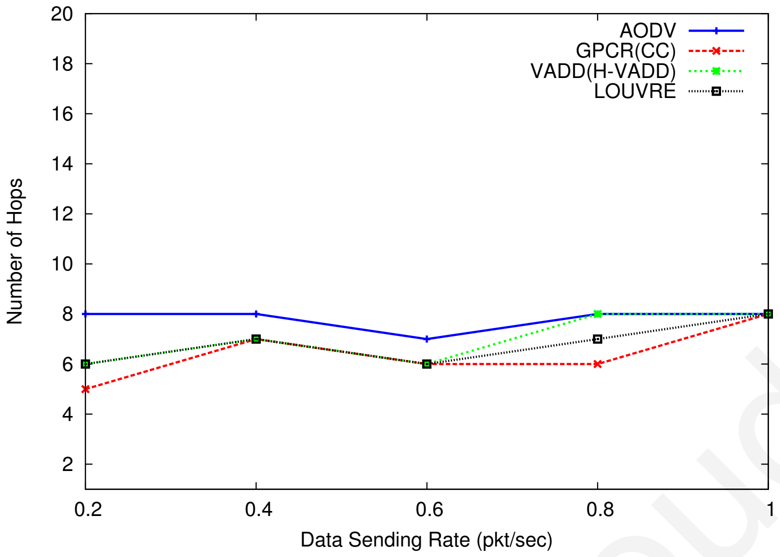


Figure 10.3: Number of Hops as a function of packet sending rate - Typical scenario for parking place reservation

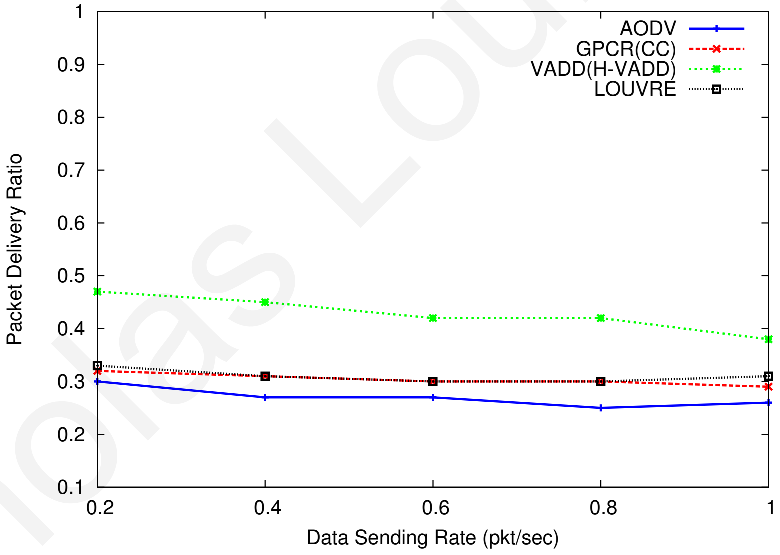


Figure 10.4: Packet delivery ratio as a function of packet sending rate - Typical scenario with V-Radar

extensively. This is due to the round-trip that each query has to perform; queries need to be forwarded at first from the source vehicle to the destination road - there, once a query reply has been formed it has to be routed back to the query source node. In addition a number of queries might be send towards some roads with no vehicles at all. We notice that AODV exhibits the lowest packet delivery ratio among all protocols and this is accounted to its backward learning mechanism. Specifically, when an node receives a query en-route to the destination road, it creates a record to its routing table containing the address of the previous node (previous hop) from which the query arrived. Upon the query arrival at the destination road, a reply is formed and send back to the origin vehicle through the path which is formed by the previous hop records at each intermediary node. However, due to the high dynamics of the vehicular environment, the structure of inter-vehicle communication changes rapidly [97], thus the backward learning process is not efficient and results to several packets being dropped due to the expiration of backward links. On the contrary, GPCR does not maintain a routing table. Queries and query replies are routed from the origin node to the destination road and back using ad hoc, geographic greedy forwarding. As in the case of sending packets to a fixed side, GPCR can fail when the VANET becomes fragmented and greedy forwarding in no longer viable. On the contrary, VADD is able to counteract such fragmentation through its aforementioned carry-and-forward capability. However, in the case where traffic does not have a relatively stable state (as in the typical example above), VADD might underestimate in the process of selecting the best path towards the destination. This can impose additional delays on a query. In an application such as V-Radar, where vehicles would like to know the traffic conditions of the roads further ahead in a relative short amount of time, the packet delivery delay exhibited by VADD is considered to be unacceptable. As identified in [102], Traffic Flow and Enhanced Route Guidance and Navigation applications such as V-Radar, should have a maximum allowable latency of 1 second. A packet delivery delay in excess of 40sec which is evident for VADD in Figure 10.5 is not considered to be acceptable since it increases the probability of acquiring stale traffic information.

Despite the presence of V-Radar, the above two cases were performed under a simple road topology with unrealistic, low-density mobility. Since mobility is key to the

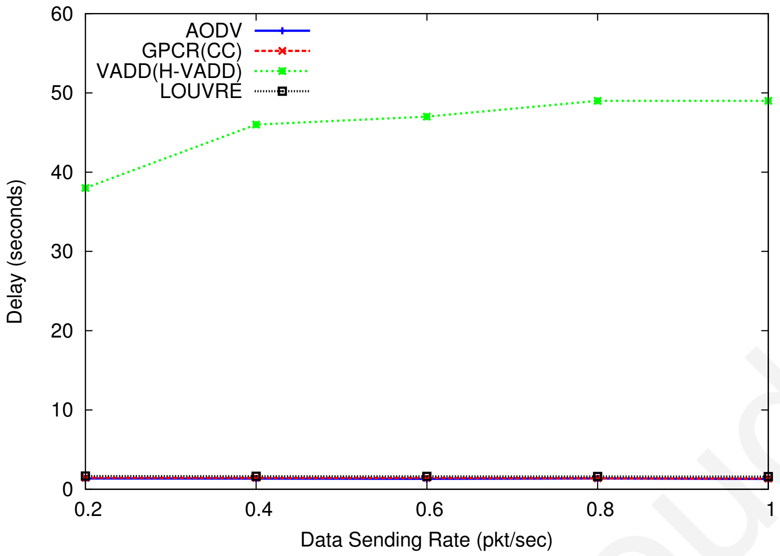


Figure 10.5: Packet delivery delay as a function of packet sending rate - Typical scenario with V-Radar

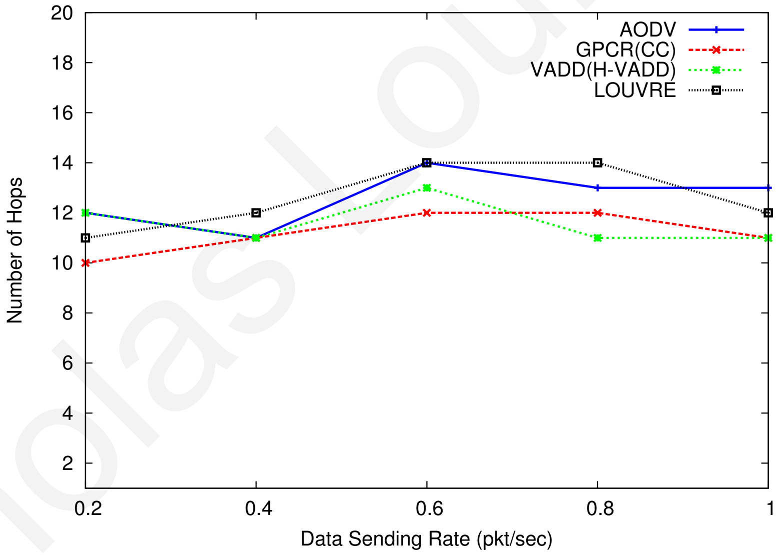


Figure 10.6: Number of hops as a function of packet sending rate - Typical scenario with V-Radar

dynamics and structure of VANETs - and consequently to the proper-function of any application and service relying on such networks - the performance of AODV, GPCR, VADD and LOUVRE stemming from the above evaluation might be considerably underestimated. The next section aims to investigate this hypothesis.

10.1.5 Large-Scale Urban Scenario

In order to evaluate AODV, GPCR, VADD and LOUVRE in a large-scale urban scenario, we again employed the TAPAS-Cologne [24] realistic mobility dataset. We utilize a reduced version of this improved dataset, which contains vehicle trips between 6:00am and 8:00am. We study the first 1000 seconds of this interval, in which a total of 4670 vehicles were emitted in the road topology.

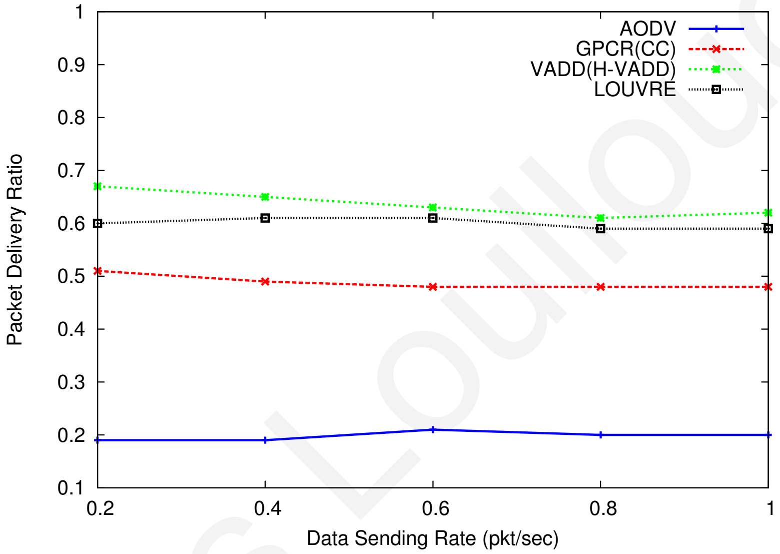


Figure 10.7: Packet delivery ratio as a function of packet sending rate - TAPAS/-Cologne

In addition all vehicles are installed with the V-Radar traffic application and generate queries as described previously in Section 10.1.4.

By looking at Figure 10.7, we immediately observe the profound effect of mobility in the performance of the studied VANET routing protocols. Because of the realistic microscopic dynamics of each individual vehicle, the correct distribution of vehicle flows in a macroscopic level and the high density in the underlying road topology, the performance of GPCR, VADD and LOUVRE improves significantly allowing traffic queries and their respective replies to be routed more efficiently and effectively. Even for a non-carry and forward protocol such as LOUVRE, realistic mobility al-

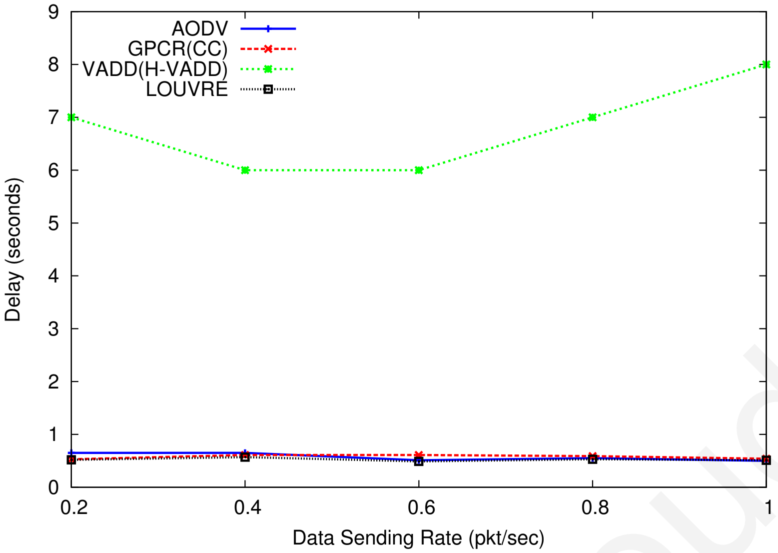


Figure 10.8: Packet delivery delay as a function of packet sending rate - TAPAS/-Cologne

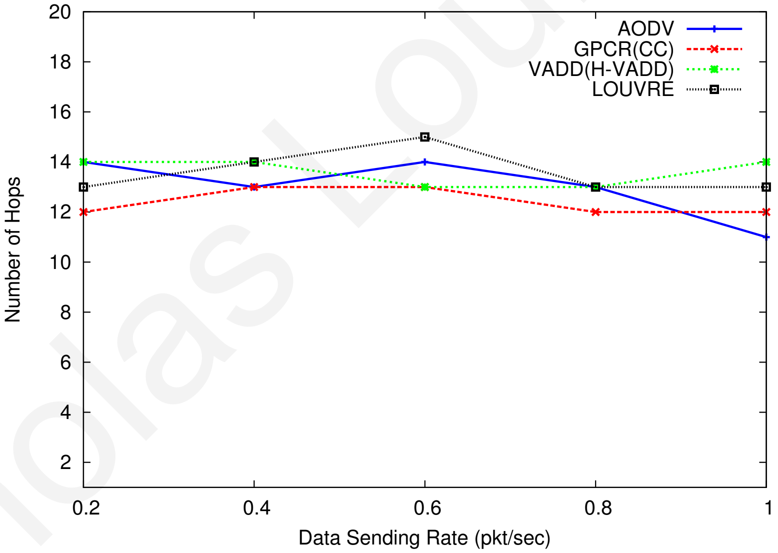


Figure 10.9: Number of hops as a function of packet sending rate - TAPAS/Cologne

lows packet delivery up to approximately 60% in comparison to the 35% in the typical scenario with V-Radar. Since traffic conditions on the TAPAS-Cologne scenario do not change as rapidly as those in the typical scenario (i.e. traffic is stabilized), the overlay network in LOUVRE is able to maintain a global vision of the density distributions on roads, and thus it can route information more efficiently than GPCR.

Since the latter is not aware of any important information concerning the underlying road topology and consequently tries to route packet from source to destination in a greedy fashion, it can encounter local maxima situations such as empty roads from which it cannot recover and eventually cause packets to be dropped. Despite that, the performance of both LOUVRE and GPCR in the TAPAS scenario outperforms their performance in the typical scenario with V-Radar. Due to the better end-to-end connectivity because of high vehicle density, greedy forwarding can achieve a higher packet delivery.

Implications: The data buffering ability of VADD, as well as the repetitive process at each encountered intersection of calculating the optimal direction a packet must take to reach its destination, allows it to avoid local maxima situations, reducing the number of packets being dropped and hence increasing the packet delivery ratio. Of course this comes as an expense to the packet delivery delay, which as it can be seen from Figure 10.8 is significantly higher than that imposed by the other protocols. GPCR, LOUVRE and even AODV are able to perform within the acceptable latency boundaries [102] for an application such a V-Radar, although we note the fact that queries are not propagated more than 2 roads ahead from the vehicles current position ($L = 2$). In situations that a vehicle would like to know the traffic conditions even further down the road-path ($L > 2$), the lack of data buffering may degrade the effectiveness of the aforementioned protocols. Furthermore, we observe from Figure 10.9 that most of the time, packets under LOUVRE require more hops from source to destination and back than GPCR and AODV. The reasoning behind this behavior, is the capacity of the LOUVRE to obtain a global view of the topology, and hence being able to route around low density areas ("voids" that cause packet drops) by introducing a few additional network hops.

10.2 Caching Dynamic Information in Vehicular Ad Hoc Networks

As seen on Section 8, VANETs are characterized by highly dynamic topologies, short-lived links, and frequent network disconnections. In addition, Traffic Information Systems (VIS) generate a substantially high number of messages in order to obtain and maintain a global or even partial view of the prevailing conditions in the vehicular environment. This overhead imposed on the VANET can lead to the saturation of the already limited network capacity, which in turn is bound to degrade the quality of VANET services (i.e: lower response times and loss of information quality) [105].

Therefore, the provision of efficient, robust, wide-area information services over vehicular ad hoc networks remains an open challenge. In this paper, we investigate the implications of caching on the efficiency of VANETs and on the quality/accuracy of VANET-based VIS. Our study is performed in the context of *VITP*, a proactive, location-oriented protocol designed for the retrieval of dynamic vehicular information over VANET [42]. In dense networks, proactive protocols like VITP can avoid saturation caused by flooding and diffusion since information is queried on demand rather than being pushed periodically during discrete time intervals. Caching may improve further the efficiency of VITP by reducing the time to serve information requests, minimizing overall network resource consumption, and improving information accessibility in the presence of mobility constraints that result to short-lived links, network disconnections, etc. However, caching may also lead to a significant degradation of VIS quality, given the highly-dynamic nature of vehicular information. In this work, we conduct an extensive exploration of the trade-off between the efficiency and the quality of VANET-based VIS services, in the presence of caching.

In contrast to earlier research efforts [50], which have shown that caching can be beneficial when vehicles are moving on unidirectional straight roads, in our work we examine the implications of caching in a number of different vehicular traffic scenarios that are more realistic and challenging. Building upon our prior work on VITP [42], we provide answers to the key question: *Does the co-existence of a proactive, location-aware, communication protocol and caching maintain acceptable levels of vehicular information quality while sustaining network performance?* Our main contributions are summarized as follows: a) We extend the VITP architecture [42] in order to support cache-based location aware services. VITP is a proactive, location-aware, application-layer, communication protocol designed to support a distributed service infrastructure over VANETs. b) Through an extensive simulation testbed, we identify the critical parameters that affect vehicular information quality in VANETs as well as demonstrate the viability and effectiveness of the cache-enabled VITP.

10.2.1 Enabling Caching support in VITP

This section presents and defines the necessary extensions to the architecture and message syntax of the Vehicular Information Transport Protocol (VITP) in order to support caching of vehicular information.

We present an extension to the architecture of VITP so that caches can be directly accessed by VITP peers. VITP peers implement the VITP protocol and operate as clients, intermediaries, or servers in a VITP-protocol interaction. The reader can find more details concerning the main design concepts of VITP architecture in [42]. We choose to extend VITP since it allows the dissemination of query messages proactively, thus providing better control in the number of messages injected in the VANET.

According to [42], a VITP transaction consists of four phases: 1) *Dispatch-query phase*: a request Q is transported through the underlying VANET toward its target area L. Q goes through a number of intermediary VANET nodes, which push the message toward its destination using geographic routing. 2) *Virtual Ad Hoc Server (VAHS)-computation phase*: the VITP request is routed between the VITP peers of the VAHS. VAHS consists of the VITP peers that contribute to computation of the reply to a VITP query. 3) *Dispatch-reply phase*: the VITP reply is geographically routed toward source region. 4) *Reply-delivery phase*: broadcasts the VITP reply to the VANET nodes of source region, so that the reply can be received by the VITP peer that originated the transaction.

In the cache-enabled architecture of VITP, the dispatch-query phase of VITP transaction is modified as follows: When a VITP peer receives a VITP query message, it checks whether this message can be served from its local cache. If there is a *cache hit*, the *last_accessed* attribute of the cached message is updated by the current time. In the context of VANETs, each object is characterized by the fields: message content and location-id. We consider a *cache hit* when a) there is a valid replica of the requested information in the cache and b) the location-ids of both requested information and its replica in the cache are in the same geographic location. Then, a VITP reply containing the cached message is generated and sent back to source node for the VITP transaction to be completed. If there is a *cache miss*, the query message is forwarded to other VITP peers with respect to the underlying geographic routing protocol. The dispatch-reply phase of the VITP transaction is also modified to support caching: when a VITP peer receives a reply message, it checks whether the reply is contained in its local cache. In case of a *hit*, the cache is updated to reflect the properties of the new message. In the case of a *miss*, the reply message is inserted into the cache and its parameters are populated accordingly. The cache system comprises a *Cache Replacement Module* (CRM), which detects stale information and evicts it from the cache. To this end, the CRM assigns a *Cache Utility Value* (CUV) to each cached message. The CRM can also be used in the unlikely case of a cache becoming full.

To support the cache operations described above, we extend the VITP message specification with a set of *cache-control* headers. These headers act as directives to

Directive	**Value**	**Description**
cacheable	Boolean	The reply can be cached.
expires	time-stamp	The time after the reply is considered expired (if cached).
private	Boolean	The cached reply can be reused only from the original peer that requested it.
public	Boolean	The cached reply can be reused from any peer.
validate-after	seconds	A peer must validate the cached reply with the target area after "n" seconds from the reply generation time.
p-validate-after	Boolean	A peer must validate the cached reply with the neighbour peers after "n" seconds since the reply generation time.
p-validate	Boolean	A peer must validate the cached reply with the neighbour peers before serving a request.
retransmit	Boolean	Serve request first from cache and retransmit it to target area also.

Table 10.2: Cache control directives used in a VITP reply message

VITP-peer caching decisions (Table 10.2). More details about the generic syntax of the VITP message along with examples can be found in [42].

10.2.2 Simulation Testbed Setup

For simulating the effects of vehicle movement, vehicular mobility traces were generated using the aid of TrafficModeller [99] and SUMO [21], a space-continuous microscopic traffic simulator. TrafficModeller is a traffic modeling and generation tool that simplifies the definition of vehicular traffic over road networks by supporting a set of high-level abstractions, such as traffic flows and population activities that induce specific traffic patterns, through an intuitive GUI. TrafficModeller allows the translation of these high-level abstractions to low-level input specification understood by SUMO. For simulating the behaviour of VITP in a VANET, ns-2 [23], was employed. To increase the level of realism and accuracy of results in the experimentation of this work, an extension to ns-2 was developed that enables the import of SUMO road networks in an appropriate data-structure prior the simulation run. This data-structure can be queried by vehicles at any instance of the simulation run-period in order to identify on which road/segment they are currently moving, thereby increasing accuracy in the resolution of location-aware queries.

10.2.3 Vehicular Mobility Generation

TrafficModeller was fed with two data sets, each one representing a region of a real city with different topographical layout. *Region 1* follows a Manhattan-like city layout where a big percentage of the road-network is comprised of long and parallel straight roads, where vehicles can accelerate to higher speeds between adjacent junctions. *Region 2* follows a more common urban layout where road segments are curved and much shorter, hence restricting vehicle acceleration. Both regions were extracted from real-world, accurate, city maps obtained from OpenStreetMap [14]. This was done in order to evaluate the efficiency of the cache-enabled VITP under real urban environments with different mobility patterns. All roads within these regions have a speed limit of 13.89 m/s. The number of junctions per squared kilometer with traffic lights is considerably much smaller in *Region 2* than in *Region 1.* It is quite important to evaluate the effectiveness of the cache-enabled VITP in road-networks with different distributions of traffic lights, since these have a direct influence in shaping the properties of vehicular traffic (i.e mean vehicle speed and density). Furthermore, a number of "hot-spots" within these regions were defined, in order to simulate the traffic conditions that arise when people drive from/to their workplaces, shopping malls, amusement centers etc. Vehicles is the simulation were set to have an acceleration of 4.5 m/s, deceleration of 2.6 m/s with their top speed bounded by the road speed limit. Driver imperfection factor, that is the ability of the driver to adapt to a desired speed, was set to 0.5 (where the value 1 indicates a perfect driver). Total simulation time was set to 1000s. These high-level abstractions were translated as low-level input specification and passed to SUMO. Vehicular traffic traces for 970 and 875 distinct vehicles that move in Region 1 and Region 2 respectively, were generated by SUMO. For both regions vehicles have a mean drive time of approximately 230s. The average vehicle speed over time for both regions is between 8 and 10 m/s. Vehicle average speeds have a large variance but follow a uniform distribution.

Mobility traces obtained from TrafficModeller and SUMO, were transformed to acceptable ns-2 input trace files using TraceExporter [25]. For the vehicular wireless network simulation, each vehicle is equipped with IEEE 802.11 capable communication hardware with a wireless radio coverage of 200m. In addition, the computing device of each vehicle is VITP enabled allowing it to participate in the resolution of incoming VITP requests as described in [42]. Furthermore, vehicle caches have an unlimited size.

Simulation scenarios run for a period of 1000s. Vehicles are injected in the network simulation starting from $t = 1s$ to $t = 950s$. A 200s warm up period is allowed before retrieving information used in the evaluation phase. The purpose of the warm-up phase is to allow the caches to reach some level of stability. Therefore, any event prior to the above time instance is not evaluated.

10.2.4 Application Scenarios and Query Generation

For the evaluation of the cache-based VITP for the exchange of different vehicular information under urban environments we have set up the following scenarios:

Scenario 1. Each vehicle is aware of the road-network topology through on-board digital maps and its current location through GPS. Since travel times are heavily influenced by the prevailing traffic conditions, vehicles would like to identify the road-paths towards their destination that if followed will result in reducing the travel-time. To discover the aforementioned conditions we introduce a query scheme implementing a “forward-scan-radar” traffic information system. According to this scheme, we assuming that a vehicle V turns into some *Road A* and wants to follow the fastest route to its destination D. Through its knowledge of the road network, V can calculate all possible road-paths connecting its current position to D. Upon entering *Road A* it issues *LookAhead (L)* queries, investigating dynamically the conditions along all these possible road paths. *LookAhead (L)* queries are propagated to a certain depth in the road-path and obtain the traffic conditions of the roads up to the specified depth.

Scenario 2.This scenario follows the paradigm of the first scenario. Here, vehicles issue *LookAhead* queries in the possible road-paths towards their destination, with the exception being that unscheduled events do take place (e.g., vehicle break-downs). Such events block roads in those paths and influence the normal traffic flow by causing following vehicles to slow-down or even stop. In case that vehicles stop there is a high probability of congestion build up which is further increased if the event has taken place on roads that are small in length and do not have alternative exits. In this scenario several vehicle break-downs throughout the network are simulated by vehicles that stop abruptly in the middle of the road for 100s and then resume their trip to their destination.

Scenario 3. We assume that vehicles would like to discover the availability of road-side facilities such as parking places, gas stations and restaurants. A stationary Road Side Unit (RSU) is responsible for each facility and at certain time intervals it broadcasts information concerning the facility (i.e free parking space availability). For our simulations, 5 RSUs were placed randomly on the road-network and broadcast information about their facility every 60 seconds. Vehicles are divided in equal groups with common facility interests and throughout the simulations randomly generate queries to identify the location of such RSUs and obtain the broadcasted information.

In all the above scenarios, VITP queries are issued with a default $ReturnCondition = 5$. A Return Condition as specified in [42] determines the sampling size of the requested information a query must obtain before a VITP reply can be generated and dispatched back to the originator of the request. For the first two scenarios, each vehicle issues queries with $LookAhead = 2$, to road paths from its current position to the its destination. The values for the above parameters were selected by sampling the parameter space having in mind that a high *LookAhead* value can cause the saturation of wireless network bandwidth thereby causing a significant amount of query drops and on the other hand a high $ReturnCondition$ value increases vehicular information accuracy. This resulted in the generation of a total of 29263 queries for *Region 1* and 16557 queries for *Region 2*. For Scenario 3, 12231 and 7562 queries were generated for *Region 1* and *Region 2* respectively.

For all the scenarios the existence of an underlying greedy geographic routing layer that forwards VITP messages from the source node towards the destination area is assumed. After each issued VITP query is satisfied by a VAHS, the generated reply is again routed geographically towards the source node. Each reply is generated with *cache-control* header = *[cache-control: cacheable, expires=t, public]*. Consequently, the information contained in each VITP reply message is cached to all intermediary nodes on its way towards the source node. Furthermore, we consider that all nodes use a *TTL-based cache replacement policy*. According to this policy, messages are removed from the cache as soon as their *TTL* (Time to Live) value expires. A *TTL* value specifies the maximum time for which a cached copy should be considered valid. Finally, Table 10.3 presents an overview of the simulation setup parameters.

10.2.5 Evaluation

To describe the performance of cache-enabled VITP in inter-vehicular networks, we employ the following metrics which are considered to be the most indicative:

- **Query Recall:** *the number of replies received while issuing queries towards a specific location of interest, over the number of replies that should have been*

Parameter	**Value**
Simulation Time	1000s
Simulation Area size	$3200m * 6800m$ (Region 1) $1600m * 1400m$ (Region 2)
Number of VITP enabled Vehicles	970 (Region 1) 875 (Region 2)
Mean Vehicle Drive Time	230 s
Wireless Coverage Area	200m
VITP Query Return Condition	5
Cache TTL values	0s to 200s
Inter-Query Issue Time	On new road-segment entry
Look-Ahead Value	2
Number of RSUs	5
RSU Broadcast Time interval	60s
Vehicle Break-Down Duration	100s

Table 10.3: Caching - Simulation Setup Parameters

received from that location. Query recall equal to 1 indicates that the protocol has managed to obtain all the requested information, while 0 denotes otherwise. For example, assume 3 locations of interest $D1$, $D2$, $D3$ and a vehicle V that generates 6 queries in total: 2 for each of the above locations. $D1$ and $D2$ have sufficient vehicles willing to resolve incoming requests and no vehicles are present at $D3$; therefore the number of replies that *should be* received from V are 4 (2 each from $D1$ and $D2$). Given that the 2 queries for $D1$ were successfully resolved, 1 query for $D2$ failed to obtain the requested information, the remaining query was dropped due to network problems, and all the queries for $D3$ were not resolved, the query recall is: $2/4 = 0.5$.

- **Response Time:** *is the average Round Trip Time (RTT) of a successful VITP transaction.* It measures the elapsed time between the time at which a query is generated and injected in the network and the time at which the corresponding reply was received at its originator. The elapsed time takes into account both VITP processing and message propagation time.

- **Information Accuracy:** *measures how close the received value describing some vehicular information is to the actual value at the location of interest.* For example assuming that we examine the average speed of vehicles on a given location, then the information accuracy metric measures how close the estimated average speed is (calculated usually by a subset of the available vehicles in the target segment), to the actual average speed in the region of interest (calculated by considering all present vehicles in the same area). If we examine the road-side facilities information, then the information accuracy metric measures how close the value retrieved from a vehicle cache is to the value broadcasted by the RSU. This metric is an indication about the quality of information that is disseminated through a vehicular network. An accuracy value equal to 1 indicates high quality information where the received information is exactly the same as the actual information.

- **Number of Exchanged Messages:** *is defined as the total number of exchanged messages, including geographic routing messages and VITP query resolution messages throughout the whole simulation period.* This metric indicates the performance improvements achieved in the reduction of network overhead through the utilization of caching in VITP.

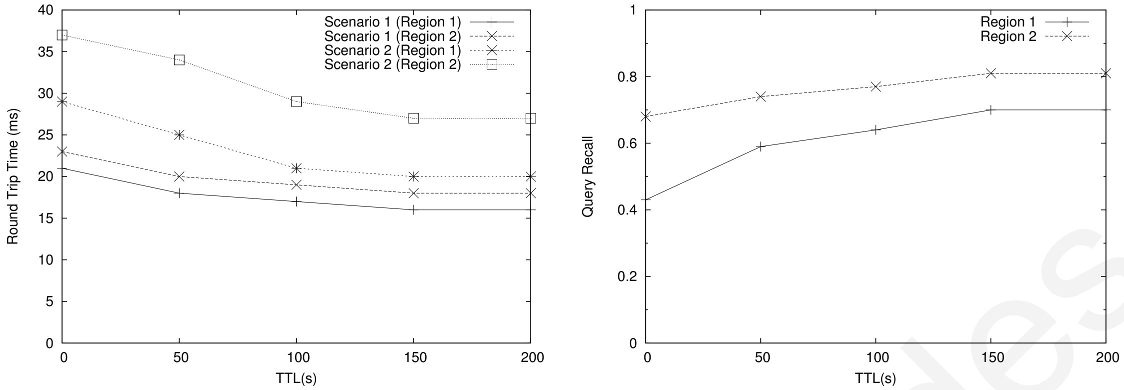
10.2.6 Caching Evaluation - Querying Road Traffic Conditions

For the three aforementioned scenarios we evaluate the performance of the cache-based VITP on both regions and under different TTL values assigned to traffic query replies. A $VITP_{TTL=0}$ emulates the original VITP where information caching is not supported on VITP peers. The maximum value of $VITP_{TTL=200}$ denotes that traffic information is cached on VITP peers for the whole drive time duration and this value was selected since it approximates the average run-time of vehicles in all the scenarios.

We begin our evaluation by investigating the performance of the cache-based VITP for Scenario 1 and Scenario 2. We measure the response time of the *LookAhead* traffic queries by varying the TTL values of cached reply messages. According to TTL-based replacement policy, when the TTL of a message expires, this cached message is discarded by all vehicles' caches. The results are reported in Figure 10.10(a) where the x-axis represents the TTL values in seconds and the y-axis represents the response time in milliseconds.

We observe that the average response time for Scenario 1 in both regions decreases with the increments in TTL. It is evident that longer TTL's result to a better diffusion of information throughout the vehicles' caches and, consequently, to an increased probability that the requested information is found nearby the requesting vehicle. Queries exhibit the lowest response time when $VITP_{TTL=200}$, with an improvement of 24% for Region 1 and 22% for Region 2 in comparison to the original VITP (no caching). It is interesting to observe that the RTT remains constant when $VITP_{TTL\geq150}$ for all the scenarios examined in this section. This occurs because vehicle queries present high temporal locality of reference. As we referred in section 10.2.2, all vehicles in the simulation issue VITP queries in the road-paths towards their destination area in order to discover the prevailing traffic conditions. Since it is not possible to change the mobility of vehicle at runtime and alter their initial route chosen by SUMO, the queries remain constant over the whole simulation time.

For Scenario 2, the average RTT for all TTL values in both regions is higher than in the case of Scenario 1. Unscheduled events like vehicle break-downs cause congestion which increases the geographic distance between querying vehicles and the query target location. Remember that when $VITP_{TTL=0}$ vehicles obtain the information only from the target location. Due to the use of geographic routing, queries have to traverse a greater number of hops in order to overcome the break-down and reach the target location. This increase in the number of hops consequently leads to an

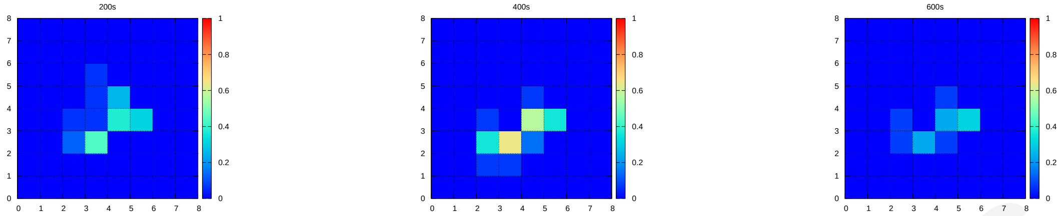


(a) (b)

Figure 10.10: (a) Query Response Time (RTT) and (b) Query Recall over TTL

increase in query round trip time. When caching is enabled, the lowest response time is encountered for $VITP_{TTL=200}$ where there is an improvement of 31% for Region 1 and 27% for Region 2.

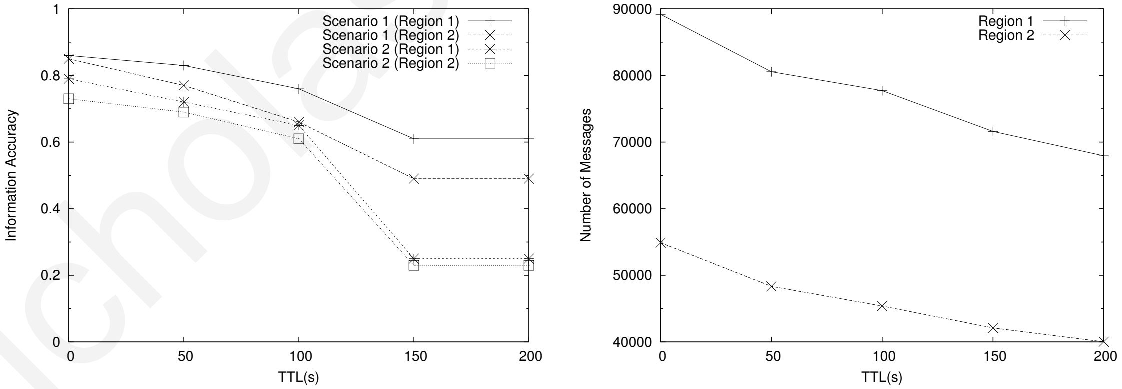
Figure 10.10(b) denotes the traffic information query recall for Scenario 1 under both regions, where the x-axis represents the TTL values and the y-axis represents the query recall percentage. The general trend is that increasing the TTL value increases the query recall for both Region 1 and Region 2. The lowest recall for both regions is when $VITP_{TTL=0}$, where no information is cached in the network and a reply to a given traffic information query can be answered only from the source location. On the other hand, the best recall is achieved when $VITP_{TTL\geq 150}$ with 70% and 81% for Region 1 and Region 2 respectively. As the TTL value increases, the number of replicated information in the network increases and a query can be served not only at its target location but also from information stored in other vehicles cache. This allows for vehicles to obtain the desired information faster and in addition reduces the amount of queries that should be re-generated in order to obtain information for the target locations of previous unresolved queries. The above observation is also reflected by examining the number and geographic distribution of replicated objects in the network (Figure 10.12). Each sub-figure illustrates a combined snap-shot of



(a) (b) (c)

Figure 10.11: Geographic distribution of Replicas in respect to simulation time

the road and wireless network. The x and y-axis denote each zone boundary in the road network while the z-axis (on the right) denotes the percentage of information completeness at each zone. Information completeness is calculated by taking into account all the distinct information replicas in a zone to the total distinct information in the VANET. It is worth noting here that replication, RTT and information accuracy are interrelated. Specifically, we observe that the pattern of dependence among them follows the rule that the increased replication results in reduced response times but also results in reduced information accuracy.



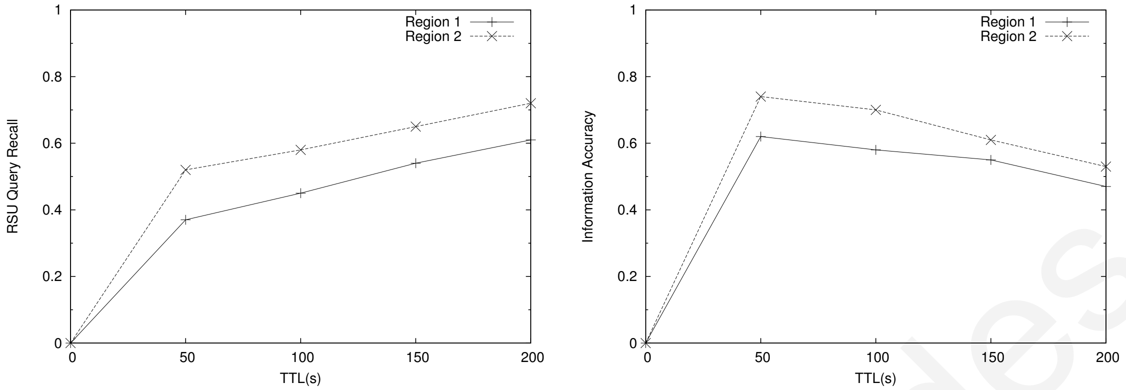
(a) (b)

Figure 10.12: (a) Information Accuracy and (b) Number of Messages over TTL

The trend of traffic information query accuracy for both scenarios is given on on Figure. 10.12(a). The x-axis represents the *TTL* values, whereas, the y-axis represents the accuracy of the information that was retrieved through the VITP queries.

For Scenario 1, the highest accuracy 86% and 85% for Region 1 and Region 2 respectively, is obtained while traffic information is retrieved directly from the query target location. Information accuracy drops as the value of *TTL* increases denoting that there is high probability that VITP queries will be answered from other vehicle's cache instead of the information source location. The drop in information accuracy is expected since cached information does not accurately reflect the traffic conditions on the queried roads. Despite this, *the cache-based VITP manages to maintain a high level of information accuracy (83% and 77% for Region 1 and Region 2 respectively) when* $VITP_{TTL=50}$*, while at the same as seen from Figure. 10.12(b)(b) can reduce the number of messages injected in the network by 11% and 12% respectively.* The noticeable difference in the rate of change of information accuracy between the two regions is due to the difference in the road-network layout. The road-network of Region 1 is comprised of several straight roads that enable vehicles to maintain a relative constant speed over a larger period of time. Therefore, in the absence of unexpected traffic events such as vehicle break-downs, cached information can maintain and report a relatively "better" representation of the prevailing road conditions of Region 1 in contrast to Region 2. The general observation here is that the TTL value of the cached information is directly influenced by the rate of change of traffic information in the roads to be queried. This rate of change is influenced by the road length, vehicle density and existence of traffic lights. Therefore, caching policies should be adaptive and take into consideration these factors in order to adjust TTL values of traffic information. In addition the results depicted on Figure 10.12(a)(a) denote that information accuracy is heavily influenced by the presence of vehicle break-downs (Scenario 2) in the road-network. For $VITP_{TTL=0}$ queries manage to capture the conditions of the road-network with relatively high accuracy (79% and 73% of the actual values for Region 1 and Region 2), but as seen previously on Figure 10.10(a), with a notably higher round-trip time. As the *TTL* value increases, there is an extensive drop in information accuracy and this can be justified as follows. For low TTL values ($VITP_{TTL<=100}$), the accuracy levels remain close to the ones in the original VITP. This is due to the fact that the specific TTL time interval is smaller than the duration of instability in the road-network (about 100s) caused by vehicle break-downs. Although in this interval several queries will be resolved from vehicles caches, a high percentage of the remaining queries will be resolved at the query target location thus obtaining a more accurate view of the prevailing road conditions. For $VITP_{TTL>100}$, the majority of queries are being resolved from information in a vehicle's cache. The low information accuracy provides evidence that cached information does not reflect correctly the conditions in the road-network.

Finally, the right diagram on Figure 10.12(b) depicts the total number of exchanged VITP messages for the resolution of traffic information queries with respect to TTL. This figure illustrates the reduction in the network overhead achieved through the cache-based VITP. We intentionally skip the evaluation for Scenario 2 in this figure, since the observed reduction of messages that would otherwise be present, would be misleading to the reader. The reduced number of messages would not be an aftermath of caching but simply because vehicles do to complete their whole routes because of break-downs and thereby issue less *LookAhead* queries. In this figure, the x-axis represents the TTL values, whereas, the y-axis represents the total number of VITP messages exchanged over the whole simulation period. We observe that caching in VANETs decreases the traffic in the network about 21% for Region 1 and about 27% for Region 2. This reduction in the number of exchanged messages results in increasing the network's reliability; fewer exchanged messages lead to fewer network failures.



(a) (b)

Figure 10.13: Caching implications on RSU discovery.

10.2.7 Caching Evaluation - Querying Road-Side Facilities Availability

The left diagram on Figure 10.13 depicts the query recall of VITP messages for the discovery of road-side facilities. From the results, it is evident that in the lack of any information diffusion mechanism, as is the case of the original VITP ($VITP_{TTL=0}$), there is a very low probability that vehicles in the simulation will locate and retrieve information broadcasted by any RSU. In order for a vehicle to retrieve such information it is necessary for it to be located within the wireless coverage area of the RSU during a specific broadcast interval and in addition to be interested for the specific information broadcasted. On the other hand, we observe that as the TTL value increases ($VITP_{TTL>0}$), the query recall increases, indicating that the utilization of caching indeed allows the diffusion of information broadcasted by the RSUs and consequently increases the probability that road-side facility related queries will be successfully answered from other vehicles' caches. Even for low TTL values ($VITP_{TTL=50}$), query recall up to 52% is achieved while broadcasted information is captured with an accuracy up to 74%. The best-case scenario for query recall is when ($VITP_{TTL=200}$), meaning that vehicles cache information broadcasted by road-side units for all their drive time duration. On the other hand, as the right diagram on Figure 10.13 denotes, increasing TTL values has a negative effect on the accuracy of the information received by querying vehicles. While ($VITP_{TTL=200}$) might give 61% and 72% query recall for Region 1 and Region 2 respectively, it only manages to reflect 47% and 53% of the information accuracy.

Chapter 11

Conclusion and Discussions

11.1 VANET Evolutionary Dynamics through Complex Network Science

Here we performed a deep and extensive analysis of the spatio-temporal characteristics of the VSN communication graph from the perspective of Complex Network theory. Particularly, we explore the dynamics of VSN in urban environments and investigate their impact in the design and implementation of dedicated routing protocols. The findings provide meaningful insights on the overall design of VSN architectures and not only to information routing. By exploiting the laws governing the networking shape of vehicular connectivity, such distributed architectures can be efficiently designed in order to improve the acquisition, storage and dissemination of heterogeneous sensory information in urban environments.

11.2 Urban Traffic Sensing and Acquisition

Here we introduced *V-Radar*, a query protocol for retrieving vehicular traffic information using V2V communications. The advantage of V-Radar over related works is that it enables the querying and acquisition of traffic information along a composite road-path, starting from a vehicle's current position towards its final destination. Specifically, V-Radar is able to query not only the initally selected road-path, but also a number of alternate paths that lead to the vehicle's destination. This allows the driver or the in-vehicle navigation system to establish a more broad and complete view of the traffic conditions that will be encountered further ahead. Such knowledge can be extremely valuable in the process of calculating a more optimal route to the destination in terms of travel time. Furthermore, this work proposes a number of components for V-Radar that can work in tandem so as to maximize the number or road-paths monitored, whilst keeping wireless transmissions and bandwidth utilization at the minimum. is its ability is its ability to monitor using location-dependant queries the traffic conditions in a number of road-paths from a vehicle's current location towards its final destination. The evaluation of V-Radar using a large-scale realistic testbed indicates a significant improvements over other related and higly-cited traffic information dissemination schemes.

11.3 Traffic Information Diffusion using Reactive

Here we argued that the design of VANET routing protocols should follow a realistic and complete evaluation in order to increase the possibility of identifying problems as well as implications that need to be considered and addressed to achieve optimal performance. We support our argument by evaluating the performance of three widely acknowledged VANET routing protocols GPCR, VADD and LOUVRE in a large-scale urban environment with realistic vehicle mobility and under network traffic generated from a novel traffic query application, called V-Radar. To our knowledge this is the first performance evaluation of these three VANET-dedicated protocols. We compare the results of our evaluation approach against the results obtained by adhering to approaches typically used in the literature. Although typical scenarios such as the above and the ones used in the literature might provide indications that a particular routing protocol has an acceptable performance, they are rarely realistic. Indeed, through such a scenario we observed that carry-and-forward techniques such as the one employed by VADD perform better than greedy forward protocols in terms of packet delivery. However, by simply introducing an application such as V-Radar that requires realistic mobility and imposes delay constraints on data delivery, carry-and-forward protocols fail to perform within acceptable limits. On the other hand, greedy forwarding routing protocols such as GPCR prove that they are able satisfy low delay requirements, however they still fail in low vehicle densities due to network fragmentation. Through, the large-scale realistic scenario, we observed that both geographic and overlay routing protocols such as GPCR and LOUVRE can perform close to carry-and-forward protocols in terms of packet delivery, while still being able to satisfy the constraints of non-delay tolerant applications. From the findings in this work, one can easily acknowledge the benefits of uptaking such a thorough and realistic approach in the performance evaluation of VANET protocols. Therefore, as future work, we plan on evaluating the performance of additional routing protocols and under a complete set of realistic safety and entertainment applications.

In addition, we evaluate the performance of TIS when utilizing caching techniques to minimize the network overhead imposed by such services. In particular, we explore the utilization of caching in VANETs by extending the architecture of VITP in order to support cache-based location aware services. We extend the syntax of VITP messages by proposing a set of cache-control headers which act as directives to VITP-peer caching decisions. Extensive simulations have been conducted to explore the feasibility of caching in VITP and analyze its performance in large-scale vehicular networks, under realistic urban environments. Simulation results have shown that the use of a TTL-based cache replacement policy in urban environments, can achieve significant improvements under both normal traffic conditions and unscheduled traffic events. In addition, results have shown that caching allows vehicles to locate and collect information from fixed RSUs that otherwise would be impossible to do so. Moreover, simulation results have also shown that the utilization of caching in VITP reduces significantly the overhead imposed on the network by minimizing the total number of exchanged messages among vehicles while requesting time-sensitive information. Ultimately, this reduces the probability of failures due to network congestion.

Chapter 12

Future Work

Although our work in Chapter 8 has provided us valuable insights into the spatio-temporal dynamics of V2X communication, at the same time is surfaced the tendency of the network to *self-destruct*. The extreme variability present in several of the measures studied, as well as the *"butterfly effects"* caused by even the slightest movement of a handful of vehicles make the prediction of future network phenomena, very hard.

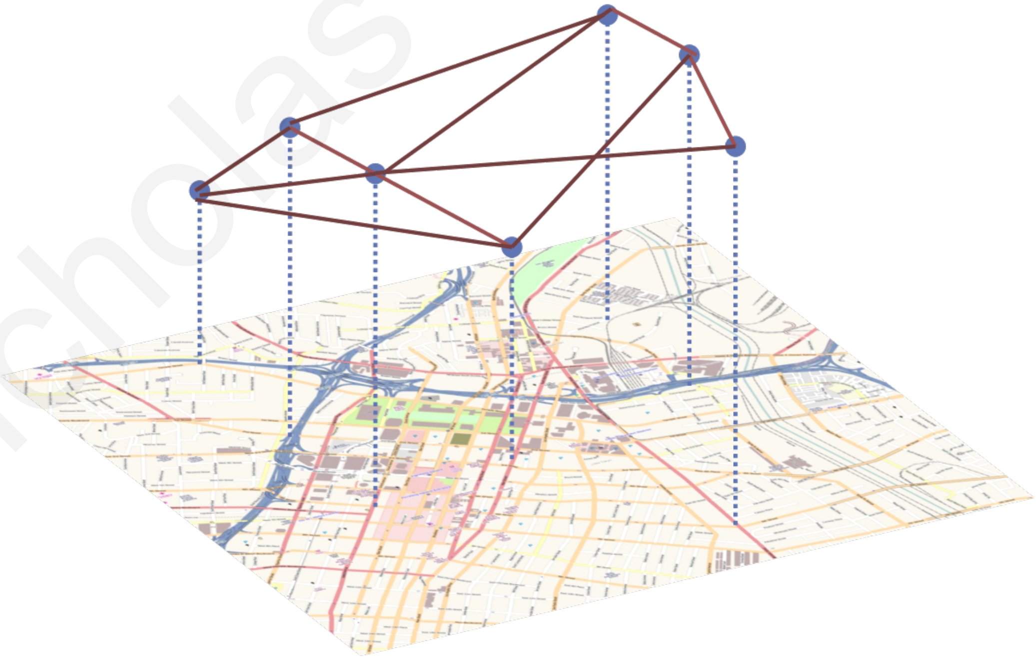


Figure 12.1: An Overlay Approach in the Dynamics of V2X Communication

As future work we have already started on studying the VANET dynamics from a number of different perspectives. The first approach aims to minimize the evident noise in the studied measures that induce the observed variability. This will ultimately improve the possibilities of forecasting the networks' evolution, even in the short-term time. For that reason we are approaching the graph from a higher level or what we term as *Overlay viewpoint*, wherein we change the connectivity model as follows:

Considering the System Model in Chapter 3, we define $TD_i(t)$ as the traffic density of road i at time t. Similarly, TD_{THLD} is defined as the traffic density threshold above which any road in the underlying network is considered as connected. A connected road is the one on which the density and distribution of V2X-enabled vehicles is such that allow information to flow freely between it's two endpoints, through the wireless medium. For each road i where $TD_i(t) \geq TD_{THLD}$, we create a respective vertex $u_i(t)$ on the graph $G(t)$. The set of edges $E(t) = \{e_{ij}\}$, correspond to the communication links among two connected roads i and j, at time t. An edge $e_{ij}(t)$ exists, if u_i is adjacent with u_j, or their geographical distance $D_{i,j} \leq R_{NLOS}$ and $i \neq j$. Effectively, as Figure 12.1 shows, $G(t)$ corresponds to an undirected graph that models the overlay network emerging from the synapses among connected roads in an urban environment.

The second approach targets the improvement of the results obtained from the current V2X modeling methodology, by substitution of contact sequences with *interval graphs*. From the literature [60, 61] it is observed that contact sequences can be biased against temporal evolving networks, since they might not cover all aspects of the temporal structure of contact patterns. As Figure 12.2 depicts, temporal networks need not to be transitive and static networks (as those used by contact sequences) do not consider transitivity, hence lacking the ability to clearly identify any information spreading dynamics.

This approach will allow us to study even more interestic metrics such as:

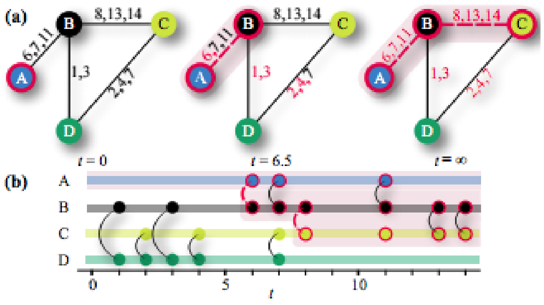


Figure 12.2: Interval Graphs[1].

- Reachability Ratio
- Maximum allowed delay of time-respective paths
- Temporal Diameter and Temporal Node Importance
- Information Propagation Latency
- ...

ultimately gaining more usefull insights to the temporal connectivity of the inter-vehicle communication that could potentially be utilized by the academia and the industry to advance the state-of-the-art in ITS solutions.

---

[1] Image taken from [60].

Bibliography

[1] Car-2-Car Communication Consortium. http://www.car-to-car.org/ (accessed August 2012).

[2] Cooperative Vehicle-Infrastructure Systems. http://www.cvisproject.org/ (accessed August 2012).

[3] Gephi - The Open Graph Viz Platform. http://gephi.github.io/.

[4] GEXF - Graph Exchange XML Format for describing complex network structures, their associated data and dynamics. http://gexf.net/.

[5] GitHub Repository. https://github.com/loulloudesn/PhD-TestBed.

[6] GML: A portable Graph File Format. http://www.fim.uni-passau.de/fileadmin/files/lehrstuhl/brandenburg/projekte/gml/gml-technical-report.pdf.

[7] Gnuplot - A Portable Command-Line Driven Graphing Utility. http://gephi.github.io/.

[8] GraphML - A Comprehensive and Easy-to-Use File Format for Graphs. http://graphml.graphdrawing.org/.

[9] HERMES Moving Object Database from the University of Piraeus, Greece . http://infolab.cs.unipi.gr/hermes/main.html.

[10] igraph - The Network Analysis package. http://igraph.org/.

[11] Japanese Ministry of Land, Infrastructure and Transport, Road Bureau ITS. http://www.mlit.go.jp/road/ITS/ (accessed August 2012).

[12] JUNG - Java Universal Network/Graph Framework. http://jung.sourceforge.net/.

[13] Mendeley Reference Manager. https://www.mendeley.com/.

[14] OpenStreetMap. http://www.openstreetmap.org/.

[15] Pajek - Program for Large Networks Analysis. http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

[16] Partners for Advanced Transit and Highways (PATH). http://www.path.berkeley.edu/ (accessed August 2012).

[17] PReVENTive and Active Safety Applications. http://www.prevent-ip.org/(accessed August 2012).

[18] Smart City Research Group. http://www.cse.ust.hk/scrg/.

[19] Smarter Traffic. http://www.ibm.com/smarterplanet/us/en/traffic_congestion/ideas/index.ht (accessed December 2013).

[20] SNAP - Stanford Network Analysis Project. http://snap.stanford.edu/.

[21] SUMO - Simulation of Urban Mobility. http://sumo.sourceforge.net/.

[22] The CabSpotting Project. http://cabspotting.org.

[23] The Network Simulator (ns-3). http://www.isi.edu/nsnam/ns/.

[24] The TAPAS-Cologne Project. http://sourceforge.net/apps/mediawiki/sumo/index.php?title=Data/Scenarios/TAPASCologne (accessed December 2013).

[25] TraceExporter. http://www.auto-nomos.de/.

[26] Traffic Data Branch, California Department of Transportation. http://traffic-counts.dot.ca.gov/.

[27] U.S. Census Bureau, TIGER system. http://www.census.gov/geo/www/tiger/.

[28] Veins - Vehicles in Network Simulation. http://veins.car2x.org/.

[29] Wigle.net - Wireless Geographic Logging Engine. http://wigle.net/.

[30] Silvia Ancona, Razvan Stanica, and Marco Fiore. Performance boundaries of massive Floating Car Data Offloading. In *2014 11th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, pages 89–96. IEEE, April 2014.

[31] Nathan Balon and Jinhua Guo. Increasing Broadcast Reliability in Vehicular Ad Hoc Networks. In *Proceedings VANET '06*, pages 104–105, Los Angeles, CA, USA, 2006. ACM.

[32] James Bernsen and D. Manivannan. Unicast Routing Protocols for Vehicular Ad Hoc Networks: A Critical Comparison and Classification. *Pervasive and Mobile Computing*, 5(1):1–18, 2009.

[33] Burton H. Bloom. Space/Time Trade-offs in Hash Coding with Allowable Errors. *Communications ACM*, 13(7):422–426, July 1970.

[34] J. Burgess, B. Gallagher, D. Jensen, and B. N. Levine. MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks. In *Proceedings INFOCOM 2006.*, pages 1 –11, April 2006.

[35] A. Chaintreau, A. Mtibaa, L. Massoulie, and C. Diot. The diameter of opportunistic mobile networks. In *Proceedings ACM CoNEXT*, 2007.

[36] Bin Bin Chen and Mun Choon Chan. MobTorrent: A Framework for Mobile Internet Access from Vehicles. In *Proceedings of the 28th IEEE International Conference on Computer Communications, IEEE INFOCOM 2009*, pages 1404–1412. IEEE Computer Society, April 2009.

[37] Wai Chen, Ratul K. Guha, Taek Jin Kwon, John Lee, and Yuan-Ying Hsu. A Survey and Challenges in Routing and Data Dissemination in Vehicular Ad Hoc Networks. *Wireless Communincations and Mobile Computing*, 11(7):787–795, jul 2011.

[38] Aaron Clauset, M. Newman, and Cristopher Moore. Finding Community Structure in Very Large Networks. *Physical Review E*, pages 1–6, 2004.

[39] Hugo Conceição, Michel Ferreira, and João Barros. On the Urban Connectivity of Vehicular Sensor Networks. In *Proceedings of the 4th IEEE International Conference on Distributed Computing in Sensor Systems*, DCOSS '08, pages 112–125, Berlin, Heidelberg, 2008. Springer-Verlag.

[40] Connect & Drive project (HTASD08002) . C&D WP-1 requirements document. Technical report, April 2009.

[41] Elizabeth M. Daly and Mads Haahr. Social Network Analysis for Routing in Disconnected Delay-tolerant MANETs. In *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '07, pages 32–40, New York, NY, USA, 2007. ACM.

[42] Marios D. Dikaiakos, Andreas Florides, Tamer Nadeem, and Liviu Iftode. Location-Aware Services over Vehicular Ad-Hoc Networks Using Car-to-Car Communication. *IEEE Journal on Selected Areas in Communications*, 25(8):1590–1602, 2007.

[43] Marios D. Dikaiakos, Saif Iqbal, Tamer Nadeem, and Liviu Iftode. VITP: An Information Transfer Protocol for Vehicular Computing. In *Proceedings of the 2nd ACM International Workshop on Vehicular Ad Hoc Networks*, pages 30–39, 2005.

[44] N. Dimokas, D. Katsaros, and Y. Manolopoulos. Cooperative Caching in Wireless Multimedia Sensor Networks. *ACM MONET*, 13(3–4):337–356, 2008.

[45] Felipe Domingos Da Cunha, Aline Carneiro Viana, Raquel A. F. Mini, and Antonio A. F. Loureiro. Is it Possible to Find Social Properties in Vehicular

Networks? In *The 9th IEEE Symposium on Computers and Communications (ISCC)*, Ilha Madeira, Portugal, June 2014.

[46] S. Dornbush and A. Joshi. StreetSmart Traffic: Discovering and Disseminating Automobile Congestion Using VANET's. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 11 –15, April 2007.

[47] Falko Dressler, Frank Kargl, Jorg Ott, Ozan K. Tonguz, and Lars Wischhof. Research Challenges in Inter-Vehicular Communication - Lessons of the 2010 Dagstuhl Seminar. *IEEE Communications Magazine*, 49(5):158–164, May 2011.

[48] Miad Faezipour, Mehrdad Nourani, Adnan Saeed, and Sateesh Addepalli. Progress and Challenges in Intelligent Vehicle Area Networks. *Commun. ACM*, 55(2):90–100, February 2012.

[49] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the Internet topology. In *SIGCOMM '99*, pages 251–262, Cambridge, Massachusetts, United States, 1999. ACM.

[50] M Fiore, F Mininni, C Casetti, and C.-F. Chiasserini. To Cache or Not To Cache? In *IEEE INFOCOM 2009 - The 28th Conference on Computer Communications*, pages 235–243. IEEE, April 2009.

[51] Marco Fiore, Claudio Casetti, and Carla-Fabiana Chiasserini. On-demand Content Delivery in Vehicular Wireless Networks. In *Proceedings of the 8th ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, MSWiM '05, pages 87–94, New York, NY, USA, 2005. ACM.

[52] Marco Fiore and Jerome Harri. The Networking Shape of Vehicular Mobility. In *Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, MobiHoc '08, pages 261–272, New York, NY, USA, 2008. ACM.

[53] Jun Gao, Jinsong Han, Dongqing Yang, and Tengjiao Wang. Road Network Based Adaptive Query Evaluation in VANET. In *Mobile Data Management, 2008. MDM '08. 9th International Conference on*, pages 49 –56, April 2008.

[54] Laura Garelli, Claudio Casetti, Carla-fabiana Chiasserini, and Marco Fiore. MobSampling: V2V Communications for Traffic Density Estimation. In *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, pages 1–5. IEEE, May 2011.

[55] Shahram Ghandeharizadeh, Shyam Kapadia, and Bhaskar Krishnamachari. An Evaluation of Availability Latency in Carrier-based Vehicular Ad Hoc Networks. In *Proceedings MobiDE '06*, pages 75–82, Chicago, Illinois, USA, 2006. ACM.

[56] Alan G. Hamilton. Gaussian Elimination. In *A First Course in Linear Algebra*, pages 1–10. Cambridge University Press, 1987. Cambridge Books Online.

[57] J. Harri, C. Bonnet, and F. Filali. Kinetic graphs: A framework for capturing the dynamics of mobile structures in MANET. In *Proceedings ACM PM2HW2N*, pages 88–91, 2007.

[58] Jerome Harri, Marco Fiore, Fethi Filali, and Christian Bonnet. Vehicular Mobility Simulation with VanetMobiSim. *Transactions of The Society for Modeling and Simulation*, September 2009.

[59] Ivan Wang-Hei Ho, Kin K. Leung, and John W. Polak. Stochastic Model and Connectivity Dynamics for VANETs in Signalized Road Systems. *IEEE/ACM Trans. Netw.*, 19:195–208, February 2011.

[60] P. Holme and J. Saramaki. Temporal Networks. *CoRR*, arxiv.org/abs/1108.1780, 2011.

[61] Petter Holme. Network Reachability of Real-World Contact Sequences. *Physical Review E*, page 46119, 2005.

[62] Bret Hull, Vladimir Bychkovsky, Yang Zhang, Kevin Chen, Michel Goraczko, Allen Miu, Eugene Shih, Hari Balakrishnan, and Samuel Madden. CarTel: A Distributed Mobile Sensor Computing System. In *Proceedings of the 4th international conference on Embedded networked sensor systems*, SenSys '06, pages 125–138, New York, NY, USA, 2006. ACM.

[63] IEEE. Draft Amendment to Standard for Information Technology-Telecommunications and Information Exchange between Systems-Local and Metropolitan Area Networks-Specific Requirements – Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications-Amendment 7: Wireless Access in Vehicular Environment. *P802.11p/D3.0*, 2007.

[64] Boangoat Jarupan and Eylem Ekici. A survey of cross-layer design for vanets. *Ad Hoc Netw.*, 9:966–983, July 2011.

[65] Jaehoon Jeong, Shuo Guo, Yu Gu, Tian He, and D. Du. TBD: Trajectory-Based Data Forwarding for Light-Traffic Vehicular Networks. In *Distributed Computing Systems, 2009. ICDCS '09. 29th IEEE International Conference on*, pages 231–238, June 2009.

[66] M. Kafsi, P. Papadimitratos, O. Dousse, T. Alpcan, and J.-P. Hubaux. VANET Connectivity Analysis. In *Proceedings of the IEEE Workshop on Autonet*, December 2008.

[67] David Kempe, Jon Kleinberg, and Amit Kumar. Connectivity and Inference Problems for Temporal Networks. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing (STOC)*, pages 504–513, New York, NY, USA, 2000. ACM.

[68] Fanyu Kong and Jindong Tan. A Collaboration-Based Hybrid Vehicular Sensor Network Architecture. In *Information and Automation, 2008. ICIA 2008. International Conference on*, pages 584–589, 2008.

[69] A. Korn, A. Schubert, and A. Telcs. Lobby index in networks. *Physica A: Statistical Mechanics and its Applications*, 388(11):2221 – 2226, 2009.

[70] Amir Krifa, Chadi Barakat, and Thrasyvoulos Spyropoulos. An optimal joint scheduling and drop policy for Delay Tolerant Networks. In *WOWMOM '08: Proceedings of the 2008 International Symposium on a World of Wireless, Mobile and Multimedia Networks*, pages 1–6, Washington, DC, USA, 2008. IEEE Computer Society.

[71] Vinod Kulathumani and Yaser Fallah. VCAST: An Infrastructure-Less Vehicular Traffic Information Service with Distance-Sensitive Precision. *2012 IEEE Vehicular Technology Conference (VTC Fall)*, pages 1–5, September 2012.

[72] Leslie Lamport. Time, Clocks, and the Ordering of Events in a Distributed System. *Commun. ACM*, 21(7):558–565, July 1978.

[73] K. Lee, M. Le, J. Harri, and M. Gerla. Taking the LOUVRE Approach. *Vehicular Technology Magazine, IEEE*, 4(1):86 –92, march 2009.

[74] Kevin C. Lee, Seung-Hoon Lee, Ryan Cheung, Uichin Lee, and Mario Gerla. First Experience with CarTorrent in a Real Vehicular Ad Hoc Network Testbed. In *MOVE'07*, Anchorage, Alaska, May 2007.

[75] Uichin Lee and Mario Gerla. A survey of urban vehicular sensing platforms. *Computers Networks*, 54(4):527–544, March 2010.

[76] Uichin Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi. Dissemination and Harvesting of Urban Data Using Vehicular Sensing Platforms. *Vehicular Technology, IEEE Transactions on*, 58(2):882–901, 2009. ID: 1.

[77] Uichin Lee, Eugenio. Magistretti, Biao Zhou, Mario Gerla, Paolo Bellavista, and Antonio Corradi. Mobeyes: Smart Mobs for Urban Monitoring with Vehicular Sensor Networks. *Wireless Commun.*, 13(5):52–57, October 2006.

[78] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceeding ACM SIGKDD*, pages 177–187, 2005.

[79] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph Evolution: Densification and Shrinking Diameters. *ACM Trans. Knowl. Discov. Data*, 1(1), March 2007.

[80] Jure Leskovec, Kevin J. Lang, Anirban Dasgupta, and Michael W. Mahoney. Statistical Properties of Community Structure in Large Social and Information Networks. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 695–704, New York, NY, USA, 2008.

[81] Yun-Wei Lin, Yuh-Shyan Chen, and Sing-Ling Lee. Routing Protocols in Vehicular Ad Hoc Networks: A Survey and Future Perspectives. *Information Science and Engineering*, 26(3):913–932, 2010.

[82] Christian Lochert, Martin Mauve, Holger Fussler, and Hannes Hartenstein. Geographic Routing in City Scenarios. *SIGMOBILE Mob. Comput. Commun. Rev.*, 9:69–72, January 2005.

[83] Christian Lochert, Bjorn Scheuermann, and Martin Mauve. Probabilistic Aggregation for Data Dissemination in VANETs. In *Proceedings of the fourth ACM international workshop on Vehicular ad hoc networks*, VANET '07, pages 1–8, New York, NY, USA, 2007. ACM.

[84] Tim. Lomax and David. Schrank. 2012 Urban Mobility Report. Technical report, Texas A&M Transportation Institute, 2012.

[85] N. Loulloudes, G. Pallis, and M. D. Dikaiakos. Caching Dynamic Information in Vehicular Ad Hoc Networks. In *Proceedings of the 16th International Conference on Parallel Processing, Euro-Par 2010*, pages 516–527, 2010.

[86] N. Loulloudes, G. Pallis, and M. D. Dikaiakos. On the Performance Evaluation of VANET Routing Protocols in Large-Scale Urban Environments (Poster). In *Proceedings of the IEEE Vehicular Networking Conference, VNC 2012*, pages 129–136. IEEE Computer Society, November 2012.

[87] Tom H. Luan, Lin X. Cai, Jiming Chen, Xuemin Shen, and Fan Bai. VTube: Towards the Media Rich City Life with Autonomous Vehicular Content Distribution. In *Proceedings of the 8th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON 2011, June 27-30, 2011, Salt Lake City, UT, USA*, pages 359–367. IEEE, 2011.

[88] Francesco Malandrino, Claudio Casetti, Carla-Fabiana Chiasserini, Christoph Sommer, and Falko Dressler. Content Downloading in Vehicular Networks: Bringing parked cars into the picture. In *23rd IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2012*, pages 1534–1539, September 2012.

[89] Friedemann Mattern. Virtual Time and Global States of Distributed Systems. In *Parallel and Distributed Algorithms*, pages 215–226. North-Holland, 1988.

[90] R. Meireles, M. Ferreira, and J. Barros. Vehicular Connectivity Models: From Single-Hop Links to Large-Scale Behavior. In *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, pages 1–5, Sept 2009.

[91] M. Moe. CVIS Communication Performance Results. In *2nd ETSI TC ITS Workshop*. ETSI, February 2010.

[92] Romeu Monteiro, Susana Sargento, Wantanee Viriyasitavat, and Ozan K. Tonguz. Improving VANET Protocols via Network Science. In *2012 IEEE Vehicular Networking Conference, VNC 2012, Seoul, Korea (South), November 14-16, 2012*, pages 17–24. IEEE, 2012.

[93] Diala Naboulsi and Marco Fiore. On the Instantaneous Topology of a Large-scale Urban Vehicular Network: the Cologne Case. In *The 14th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc '13, Bangalore, India*, pages 167–176, New York, NY, USA, July 2013. ACM.

[94] Tamer Nadeem, Sasan Dashtinezhad, Chunyuan Liao, and Liviu Iftode. TrafficView: Traffic Data Dissemination Using Car-to-Car Communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 8(3):6–19, 2004.

[95] Alok Nandan, Shirshanka Das, Giovanni Pau, Mario Gerla, and M. Y. Sanadidi. Co-operative downloading in vehicular ad-hoc wireless networks. In *Proceedings of the Second Annual Conference on Wireless On-demand Network Systems and Services*, WONS '05, pages 32–41, Washington, DC, USA, 2005. IEEE Computer Society.

[96] M. Newman and M. Girvan. Finding and Evaluating Community Structure in Networks. *Physical Review E*, 69(2), 2003.

[97] G. Pallis, D. Katsaros, M.D. Dikaiakos, N. Loulloudes, and L. Tassiulas. On the Structure and Evolution of Vehicular Networks. In *Modeling, Analysis Simulation of Computer and Telecommunication Systems, 2009. MASCOTS '09. IEEE International Symposium on*, pages 1–10, Sept. 2009.

[98] S. Panichpapiboon and W. Pattaratikom. Connectivity Requirements for Self-Organizing Traffic Information Systems. *IEEE Transactions on Vehicular Technology*, 57(6):3333–3340, November 2008.

[99] L.G. Papaleondiou and M.D. Dikaiakos. TrafficModeler: A Graphical Tool for Programming Microscopic Traffic Simulators through High-Level Abstractions. In *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pages 1–5, April 2009.

[100] European Parliament. Directive 2010/40/EU of the European Parliament and of the Council of 7 July 2010, on the framework for the deployment of Intelligent Transport Systems in the field of road transport and for interfaces with other modes of transport. Technical report, European Parliament, 2010.

[101] C. Perkins, E. Belding-Royer, and S. Das. Ad hoc On-Demand Distance Vector (AODV) Routing, 2003.

[102] Vehicle Safety Communication Project. Task 3 Final Report: Identify Intelligent Vehicle Safety Applications. Technical report, U.S. Dept. of Transport - National Highway Traffic Safety Administration, March 2005.

[103] Narayanan Sadagopan, Fan Bai, Bhaskar Krishnamachari, and Ahmed Helmy. PATHS: Analysis of PATH Duration Statistics and their Impact on Reactive MANET Routing Protocols. In *Proceedings MobiHoc '03*, pages 245–256, Annapolis, Maryland, USA, 2003. ACM.

[104] Amit Kumar Saha and David B. Johnson. Modeling mobility for vehicular ad-hoc networks. In *VANET '04: Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, pages 91–92, New York, NY, USA, 2004. ACM.

[105] Bjorn Scheuermann, Christian Lochert, Jedrzej Rybicki, and Martin Mauve. A Fundamental Scalability Criterion for Data Aggregation in VANETs. In *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, MobiCom '09, pages 285–296, New York, NY, USA, 2009. ACM.

[106] E. Schoch, F. Kargl, M. Weber, and T. Leinmuller. Communication Patterns in VANETs. *Communications Magazine, IEEE*, 46(11):119–125, 2008.

[107] Ramon S. Schwartz, Anthony E. Ohazulike, Christoph Sommer, Hans Scholten, Falko Dressler, and Paul Havinga. Fair and adaptive data dissemination for Traffic Information Systems. *2012 IEEE Vehicular Networking Conference (VNC)*, pages 1–8, November 2012.

[108] Karim Seada. Insights from a Freeway Car-to-car Real-world Experiment. In *Proceedings of the Third ACM International Workshop on Wireless Network Testbeds, Experimental Evaluation and Characterization*, WiNTECH '08, pages 49–56, New York, NY, USA, 2008. ACM.

[109] Antonios Skordylis and Niki Trigoni. Delay-bounded Routing in Vehicular Ad Hoc Networks. In *Proceedings of the 9th ACM international symposium on Mobile ad hoc networking and computing*, MobiHoc '08, pages 341–350, New York, NY, USA, 2008. ACM.

[110] Christoph Sommer, Ozan Tonguz, and Falko Dressler. Traffic information systems: efficient message dissemination via adaptive beaconing. *IEEE Communications Magazine*, 49(5):173–179, May 2011.

[111] V. Srivastava, A. B. Hilal, M. S. Thompson, J. N. Chattha, A. B. MacKenzie, and L. A. DaSilva. Characterizing mobile ad hoc networks: The MANIAC challenge experiment. In *Proceedings ACM WiNTECH*, pages 65–72, 2008.

[112] Stephenson, Karen and Zelen, Marvin. Rethinking centrality: Methods and examples. *Social Networks*, 11(1):1–37, Mar 1989.

[113] Ozan K. Tonguz and Mate Boban. Multiplayer games over Vehicular Ad Hoc Networks: A new application. *Ad Hoc Netw.*, 8(5):531–543, July 2010.

[114] Yasser Toor, Paul Muhlethaler, Anis Laouiti, and Arnaud La Fortelle. Vehicle Ad Hoc networks: Applications and Related Technical Issues. *IEEE Communications Surveys & Tutorials*, 10(3):74–88, 2008.

[115] S. Uppoor and M. Fiore. Large-Scale Urban Vehicular Mobility for Networking Research. In *Vehicular Networking Conference (VNC), 2011 IEEE*, page 62, November 2011.

[116] W. Viriyasitavat, F. Bai, and O.K. Tonguz. Dynamics of Network Connectivity in Urban Vehicular Networks. *Selected Areas in Communications, IEEE Journal on*, 29(3):515 –533, March 2011.

[117] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications.* Cambridge University Press, 1994.

[118] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, (393):440–442, 1998.

[119] L. Wischoff, A. Ebner, H. Rohling, M. Lott, and R. Halfmann. SOTIS - A Self-Organizing Traffic Information System. In *Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual*, volume 4, pages 2442 – 2446, April 2003.

[120] Xianren Wu, Hamid R. Sadjadpour, and J. J. Garcia-Luna-Aceves. Link dynamics in MANETs with Restricted Node Mobility: Modeling and Applications. *Trans. Wireless. Comm.*, 8(9):4508–4517, 2009.

[121] Xue Yang, Jie Liu, Feng Zhao, and Nitin H. Vaidya. A Vehicle-to-Vehicle Communication Protocol for Cooperative Collision Warning. In *MobiQuitous*, pages 114–123, 2004.

[122] J Y Yen. Finding the K Shortest Loopless Paths in a Network. *Management Science*, 17(11):712–716, 1971.

[123] Jijun Yin, Tamer ElBatt, Gavin Yeung, Bo Ryu, Stephen Habermas, Hariharan Krishnan, and Timothy Talty. Performance Evaluation of Safety Applications over DSRC Vehicular Ad Hoc Networks. In *Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks*, VANET '04, pages 1–9, New York, NY, USA, 2004. ACM.

[124] Jijun Yin, Tamer ElBatt, Gavin Yeung, Bo Ryu, Stephen Habermas, Hariharan Krishnan, and Timothy Talty. Performance Evaluation of Safety Applications over DSRC Vehicular Ad Hoc Networks. In *Proceedings VANET '04*, pages 1–9, Philadelphia, PA, USA, 2004. ACM.

[125] Eiko Yoneki. Visualizing Communities and Centralities from Encounter Traces. In *Proceedings of the Third ACM Workshop on Challenged Networks*, CHANTS '08, pages 129–132, New York, NY, USA, 2008. ACM.

[126] J. Zhao and G. Cao. VADD: Vehicle-Assisted Data Delivery in Vehicular Ad Hoc Networks. In *INFOCOM 2006. 25th IEEE International Conference on Computer Communications. Proceedings*, pages 1–12, April 2006.

[127] Huang Zhu and Gurdip Singh. A Communication Protocol for a Vehicle Collision Warning System. In *Proceedings of the 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing*, GREENCOM-CPSCOM '10, pages 636–644, Washington, DC, USA, 2010. IEEE Computer Society.