an Efficient emergency information dissemination scheme for emerging infrastructure-based vehicular networks

Raoua Chakroun

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de Toulouse

---

Présentée et soutenue par

Raoua CHAKROUN

Le 30 novembre 2022

un schéma efficace de dissémination des informations urgentes pour les réseaux véhiculaires émergents

---

Ecole doctorale : **SYSTEMES**

Spécialité : **Informatique**

Unité de recherche :
LAAS - Laboratoire d'Analyse et d'Architecture des Systèmes

Thèse dirigée par
Thierry VILLEMUR et Slim ABDELLATIF

Jury

M. Nazim AGOULMINE, Rapporteur
M. Toufik AHMED, Rapporteur
M. Ernesto EXPOSITO, Examinateur
M. Thierry VILLEMUR, Directeur de thèse
M. Slim ABDELLATIF, Co-directeur de thèse
Mme Marie-Pierre GLEIZES, Présidente

Remerciements

Les travaux présentés dans ce manuscrit ont été réalisés au sein de l'équipe SARA (Services et Architectures pour Réseaux Avancés) du LAAS-CNRS (Laboratoire d'Analyse et d'Architectures des Systèmes). Je tiens donc à remercier Messieurs Liviu NICU (directeur du LAAS) et Khalil DRIRA pour leur accueil dans ce laboratoire. Je tiens à exprimer ma profonde reconnaissance et gratitude à mes directeurs de thèse Thierry VILLEMUR et Slim ABDELLATIF pour la confiance dont ils ont fait preuve à mon égard et surtout d'avoir accepté de travailler avec moi. Leurs divers conseils et directives m'ont permis d'accomplir ce travail avec succès. Je les remercie également pour tout ce qu'ils m'ont appris autant scientifiquement que personnellement.

Je désire en outre adresser mes sincères remerciements aux membres de jury qui ont accepté d'évaluer ce travail. Je tiens d'abord à remercier Messieurs Toufik AHMED et Nazim AGOULMINE, d'avoir accepté de rapporter ce travail et pour leurs commentaires et suggestions perspicaces. Je remercie également Madame Marie-Pierre GLEIZES (présidente du jury) et Monsieur Ernesto EXPOSITO d'avoir accepté d'examiner ce travail. Je les remercie tous pour la qualité des discussions lors de ma soutenance de thèse.

Un grand merci également à toutes les personnes que j'ai côtoyées depuis mon arrivée au LAAS, qui ont tous largement contribué à l'excellente atmosphère du laboratoire et avec qui j'ai partagé des moments inoubliables.

Enfin, je remercie ma famille et mes amis proches, pour leurs encouragements depuis le début de ma thèse, surtout mes parents pour leurs soutiens durant toutes mes années d'études, je les remercie pour les valeurs qu'ils m'ont transmis qui ont participé à faire de moi ce que je suis aujourd'hui, aucun mot ne serait suffisant pour leur exprimer ma reconnaissance et ma gratitude.

Résumé : Les applications et les technologies des véhicules connectés envisagées dans le cadre des Systèmes de Transport intelligent (ITS) doivent pouvoir répondre à une grande variété de besoins. En particulier, elles assistent les conducteurs pour anticiper les accidents potentiels et elles aident les utilisateurs et les opérateurs du système de transport à effectuer des choix intelligents, économiques voire écologiques, réduisant ainsi les retards de voyage, diminuant les embouteillages, et fournissant des services de sécurité et de divertissement aux personnes sur la route. Les besoins en communication nécessaires aux applications liées aux véhicules connectés sont supportés par des réseaux véhiculaires, qui combinent différentes technologies d'accès et d'architectures réseau.

Parmi l'ensemble des services de communication utilisables, nous nous sommes focalisés sur ceux requis par les applications de sécurité telles que la sensibilisation coopérative (par exemple, alerte en cas d'accident, etc.). Ces services nécessitent une diffusion de messages d'alerte efficace, sans perte et avec une couverture suffisamment importante. A l'origine, la dissémination des messages d'alerte était effectuée en invitant les véhicules à rediffuser à l'aveugle les messages d'alerte. Cependant, cette inondation inefficace provoque une quantité massive de retransmissions et de collisions inutiles, gaspille la bande passante, augmente le délai de diffusion et réduit au final le taux de livraison des paquets. Face à ce problème, diverses approches coordonnant la dissémination de messages d'urgence ont été proposées. Notre travail de thèse s'inscrit dans ce cadre. Nous développons une nouvelle procédure de dissémination des messages d'alerte basée sur la localisation pour les futurs réseaux véhiculaires. Trois contributions sont développées. La première définit notre nouvelle procédure de dissémination des messages d'alerte. Cette procédure se base sur la définition de zones de diffusion. Dans chaque zone, notre algorithme se charge d'élire un véhicule relai en fonction de sa localisation. La seconde contribution améliore la précédente en affinant le calcul du placement des relais et en le rendant plus dynamique. Nous développons une approche basée sur l'apprentissage par renforcement. Cette approche est capable d'ajuster le placement optimal des points/zones de rediffusion où les véhicules peuvent relayer le message afin d'atteindre les véhicules roulant dans des zones qui souffrent d'une mauvaise/absence de couverture réseau (zones grises) et donc conduire à une couverture d'information le plus optimale possible. La troisième contribution propose un service d'estimation de la qualité de liens basé sur des techniques d'apprentissage automatique (Machine Learning). Ce service vise à offrir une vision potentielle de l'état futur du réseau, ce qui permet aux fonctions de contrôle réseau de mettre à jour au besoin les points relais afin d'assurer une meilleure couverture d'information.

Mots clés : Réseaux véhiculaires, Système de Transport Intelligent (ITS), Dissémination, Application de sécurité routière, Machine Learning (ML)

Abstract: The connected vehicle applications and technologies envisaged for Intelligent Transport Systems (ITS) must be able to meet various needs. In particular, they assist drivers in anticipating potential accidents and help users and operators of the transportation system make intelligent, economical, and environmentally friendly choices, reducing travel delays, decreasing traffic congestion, and providing safety and entertainment services to people on the road. Vehicular supports communication requirements for connected vehicle applications, combining different access technologies and network architectures.

Among the set of usable communication services, we focused on those required by safety applications such as cooperative awareness (e.g., accident alert, etc.) because these services require effective, lossless broadcast and dissemination of alert messages with sufficient coverage. Historically, alert message dissemination was performed by blindly inviting vehicles to rebroadcast alert messages. However, this inefficient flooding causes massive unnecessary retransmissions and collisions, wastes bandwidth, increases broadcast delay, and ultimately reduces packet delivery rate. In response to these issues, various approaches have been proposed. Our thesis work is part of this context. We develop a new location-based alert message dissemination procedure for future vehicular networks. Three contributions are proposed. The first defines our new procedure for disseminating alert messages. This procedure is based on the rebroadcast zones definition method, where relay vehicles are selected according to their location to rebroadcast the message. The second contribution improves the previous by refining and optimizing the relay placement computation. We develop an approach based on reinforcement learning. This approach can adjust the optimal placement of the relay points/areas where vehicles can relay the message to reach vehicles driving in areas that suffer from poor/no network coverage (grey areas) and thus lead to the most optimal information coverage possible. The third contribution proposes a link quality estimation service based on Machine Learning techniques. This service aims at providing a potential vision of the future state of the network, allowing network control functions to update relay points as needed to ensure better information coverage.

Keywords: Vehicular network, Intelligent Transport System (ITS), Dissemination, Road safety application, Machine Learning (ML)

Contents

List of Figures

List of Tables

Nomenclature

AODV Ad hoc On-Demand Distance Vector

BSS Basic Service Set

C-ITS Cooperative Intelligent Transport System Intelligent Transport

CAM Cooperative Awareness Message

CSMA Carrier Sense Multiple Access

DENM Decentralized Environmental Notification Message

DSRC Dedicated Short Range Communication

EDCA Enhanced Distributed Channel Access

ITS Intelligent Transport System Intelligent Transport

MANET Mobile Ad-hoc Network

MBMS Multimedia Broadcast and Multicast Services

MDP Markov Decision Process

OFDM Orthogonal Frequency Division Multiplexing

RL Reinforcement Learning

RSU Road Side Unit

SDN Software Defined Network

SDVN Software Defined Vehicular Network

V2I Vehicle to Infrastructure

V2N Vehicle to Network

V2P Vehicle to Pedestrian

V2V Vehicle to Vehicle

VANET Vehicular Ad hoc NETwork

WAVE Wireless Access in Vehicular Environment

Chapter 1

Introduction

1.1 Context and motivation

Despite technological and strategic advances in road safety, most transport systems worldwide still suffer from serious safety and efficiency problems. For example, the last report published by the French's National Observatory of Road Safety (ONISR) [ONISR 2022] shows that in 2019, the death rate on the road in France increased for the third consecutive year, an increase that mainly affects cyclists and pedestrians, with over 9% and 19% respectively compared to 2018. Another study shows that in 2019, French motorists spent 163 hours in traffic jams in the Paris region [TOMTOM 2022]. In addition, driving's ways affect fuel consumption by up to 20% compared to 2018[Försterling 2015].

In order to overcome these issues, Intelligent Transport Systems (ITS) are developed to provide communication capabilities to vehicles and transport infrastructures. More specifically, Cooperative ITS (C-ITS) aims to enable communications between vehicles (V2V: Vehicle-to-Vehicle), vehicles, and infrastructure (V2I: Vehicle-to-Infrastructure), and more generally, vehicles and all their surroundings (V2X: Vehicles-to-everything). These systems aim to offer reliable (fewer accidents on the road), optimal (less time on the road and less polluting), and much more comfortable (passenger entertainment: multimedia, infotainment, etc.) mobility.

The combination of communication capabilities with vehicle sensors' sensing and perception capabilities opens the way to developing countless C-ITS services and use cases. A wide range of C-ITS services has been proposed in the literature, including traffic warning/information dissemination, traffic guidance, electronic toll collection, parking spot finding, etc. They are designed to support the main C-ITS objectives. For example:

- The emergency vehicle warning service aims to warn vehicles of an emergency vehicle (e.g., ambulance) by disseminating an alert message to neighboring vehicles on the road.
- The Cooperative Collision Avoidance (CCA) service allows vehicles to avoid accidents by exchanging some mobility information.
- The bird's-eye view service allows each vehicle to build a global vision of its surroundings by combining the neighboring vehicles' local images to make efficient decisions (e.g., planning future trajectories).

- The cooperative driving service (Platonning) aims to group vehicles in platoons, or road convoys, composed of several closely spaced vehicles to save fuel consumption, prevent accidents and optimize road use.

As a result, the next generation of vehicles will be connected and equipped thanks to one or more interfaces to communicate with other intelligent transportation system elements, including other vehicles, pedestrians, and infrastructure (Roadside Units (RSU), Base Stations (BS), cloud, etc.). Therefore, these communication capabilities open the opportunity for next-generation connected vehicles to have centralized network control and the availability of vehicles-to-infrastructure network connectivity supporting traffic control.

Under this assumption, several works suggest developing a hybrid multi-access vehicular network based on Software Defined Network (SDN) to take advantage of the various access networks' capacities and bring flexibility in their control and management. This flexibility is necessary to efficiently manage the network resources and provide communication services adapted to the requirements of C-ITS services [Toufga 2017, He 2016, Salahuddin 2014]. For these reasons, this work considers SDN-based vehicular networks.

On the other hand, an SDN controller would maintain visibility of the vehicular network's current state, which usually suffers from connection issues due to changing vehicle mobility and traffic conditions. Therefore, it could identify areas that suffer from a poor or no network connection (gray zones), thus proacting and warning the network entities to react in order to provide the required network connectivity with the level of performance required by the various C-ITS services.

Alert Message (AM) dissemination is a core C-ITS safety service that, upon an emergency or a risky situation (accident or vehicle breakdown), alert messages are generated for some time and spread over a geographical area of interest as fast as possible to allow other vehicles to react quickly and adequately [Salim 2007]. This service is very demanding in terms of QoS; it requires full information coverage (all vehicles in the region must be informed), a fast dissemination time, and high reliability. Historically, message alert dissemination was carried out by a Direct Short Range Communication (DSRC) based Vehicular Adhoc NETworks (VANET) operating on a dedicated communication channel by inviting vehicles to rebroadcast the alerts blindly. However, this method, known as simple flooding, causes a huge amount of unnecessary re-transmissions and collisions (known as the broadcast storm problem), thus wasting bandwidth, increasing dissemination delay, and lowering packet delivery ratio [Tseng 2002]. Many proposals from the literature tried to mitigate the broadcast storm problem by controlling the vehicles allowed to rebroadcast (using a random selection or clustering techniques to group vehicles) or reducing the contention. But, most are complex and designed for VANETs, and neither assume the presence of V2I links nor take advantage of the benefits that a global view of the network at a centralized controller can bring to the dissemination scheme.

From this perspective, this Ph.D. aims to develop a new efficient location-based

alert message dissemination scheme for future infrastructure as SDN-based vehicular networks.

1.2 Thesis contributions

This thesis addresses the following research problems.

"Q1: How to ensure efficient dissemination of alert messages through the infrastructure to avoid broadcast storm problems while providing full message coverage with low delays?"

Location-based Alert Message Dissemination (LAMD) procedure.
To answer this question, the first contribution introduces an efficient alert messages dissemination scheme that primarily exploits V2I transmissions and complements with V2V transmissions in order to reach vehicles in areas that are poorly or not covered (gray zones) by transmissions from the infrastructure. More precisely, the network controller computes rebroadcast locations (or rebroadcast zones) to ensure that V2I broadcasts and V2V rebroadcasts lead to the full coverage of an area of interest. The selection process allows vehicles to rebroadcast an alert message based on their close vicinity to rebroadcast zones, which can be placed statically (considering network planning) or dynamically. The number and the location of rebroadcast zones are inputs to the LAMD dissemination procedure.

"Q2: How do optimally place the rebroadcast zones to cover thoroughly gray zones when applying LAMD?"

Q-learning-based rebroadcast zones placement for an effective alert message dissemination.
In the second contribution, we propose a novel reinforcement learning-based method for V2V rebroadcast zone placement to enhance and ensure the dissemination procedure's efficiency. This latter consists in optimally adjusting the rebroadcast zone placement according to the evolution of gray zones caused by changing traffic conditions. Our algorithm takes as input the observed link quality, consisting of gray zone delimitations. These can be either static (based on network planning) or dynamic (based on link quality estimation methods).

"Q3: How to predict link quality in the area of interest and the delimitations of gray zones to update rebroadcast zones proactively?"

SDN-enabled machine learning road link quality estimation
The last contribution introduces an intelligent link quality estimation solution to predict gray zone positions. We propose a prediction model based on Machine Learning techniques. This model aims to estimate vehicles' Packet Reception Ratio (PRR) of RSU's messages by zone according to the current traffic conditions on the road. Then, through the estimated PRR per zone, we can identify gray zone positions.

Figure 1.1 summarizes the proposed LAMD scheme. First, an SDN controller identifies gray zones in its geographic map coverage by applying the *"Road link quality estimation"* algorithm (contribution 3). Then it uses the position of

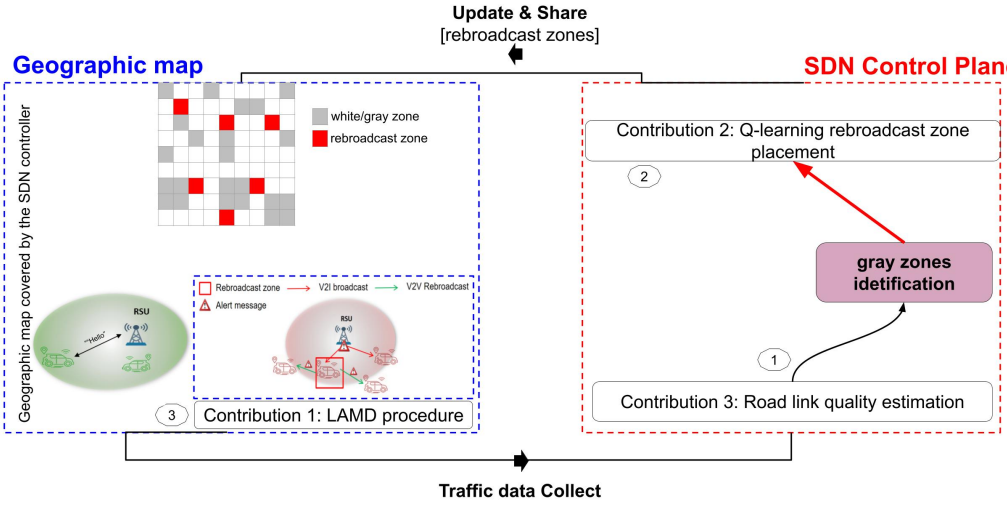


Figure 1.1: Location-based Alert Message Dissemination (LAMD) scheme

these zones as input to the *"Q-learning rebroadcast zone placement"* algorithm (contribution 2) in order to place or update the rebroadcast zones, which are shared with vehicles proactively during handover. Finally, when receiving an alert message, vehicles run the *"LAMD procedure"* (contribution 1) using rebroadcast zones as input.

1.3 Thesis organization

This thesis is composed of four chapters and is structured as follows:

Chapter 2 provides the key concepts of the C-ITS system and its services and vehicular networks and a literature overview of data dissemination to give the background information underlying this thesis work. The chapter also introduces the alert message dissemination problem and thesis assumptions to identify the thesis scope.

Chapter 3 addresses the problem of alert message dissemination. Assuming the availability of the central view of an SDN controller, we present our efficient location alert message dissemination procedure for SDVN, which collaboratively incorporates V2I links broadcasts and V2V rebroadcasts. The SDN controller drives V2V rebroadcast selection, where a simple default method is proposed to place rebroadcast zones.
First, we present the global vision of this work and its positioning compared to the scientific literature. Then, we define the key principles of our system model and the default rebroadcast zone placement. We next describe the proposed dissemination algorithm. Finally, we evaluate and discuss our simulation results.

In *Chapter 4*, a novel rebroadcast zone placement solution is presented. We propose an optimal rebroadcast zone placement method capable of adjusting their placement according to the evolution of the gray zones due to traffic fluctuations in

order to ensure more efficient dissemination and full coverage of alert messages as quickly as possible. This method is based on reinforcement learning algorithms.
First, we present the problem of rebroadcast zone placement and position our solution regarding those existing in the scientific literature. Then, we highlight the weaknesses of our default placement proposed in Chapter 2. Next, the proposed placement method has been developed. Finally, its experimental evaluation results are presented, analyzed, and compared with the default method by applying the proposed dissemination algorithm.

Chapter 5 is dedicated to the link quality estimation algorithm based on Machine Learning techniques. It aims to identify and update gray zones in response to traffic condition changes.
First, we present our proposals' quality link estimation and positioning with those existing in the scientific literature. Then, we define this model's various stages of development, from data collection through training to performance evaluation.

To conclude, the manuscript finally exposes a global vision of the work done during this thesis and the perspectives of evolution and improvements of the results.

Chapter 2

Background and Thesis Scope

2.1 Introduction

Extensive transportation expectations and advances in modern technology have led to rapid growth in the usage of vehicles in today's world [KL 2017]. However, with this rapid increase in the number of vehicles together with many additional factors, such as limited space in road infrastructure and bad driving habits, the transportation industry has been facing various problems such as an increase in traffic accidents, prolonged traffic congestion, damages to public property and human lives, inability in finding parking spots, etc. [Sudheer 2017]. In order to overcome these problems and make the transportation systems more efficient, much research is done to contribute to the evolution of vehicular networks toward an Intelligent Transport System (ITS). Their goal is to provide vehicles with ubiquitous on-road network connectivity (anytime and anywhere) with the level of performance required by various ITS services. These services are designed around the main objectives of ITS to reduce road accidents and improve the driving experience. Some services have stringent requirements considering network quality of service, mainly latency and transmission reliability.

The two leading complementary communication access technologies considered for future vehicular networks are Dedicated Short Range Communication (DSRC), based on Vehicular Ad hoc NETwork (VANET) and cellular communications. The former is foreseen for rapid dissemination and delivery of critical safety messages in a limited area of interest. At the same time, the latter is more suited for heavier and less stringent message exchanges [Xu 2017]. Also, in line with the recent trends in communication networks with the progressive adoption of network softwarization, Software Defined Networking (SDN) is considered a promising paradigm for future vehicular networks. In addition, SDN paves the way for effective network control applications to unify the control of heterogeneous multi-access vehicular networks: it can leverage the vehicular network's centralized, multi-level and multi-access view with complementary vehicles road and route traffic-related information brought from the cloud. Moreover, SDN is a key paradigm for supporting the diverse Quality of Service (QoS) expectations (some being very stringent) of C-ITS services.

In this thesis, we consider such an SDN-enabled multi-access vehicular network. Under this assumption, Vehicle to Infrastructure (V2I) wireless connectivity is nominally available, with infrequent and time-limited connectivity losses. This holds for LTE and DSRC-based access networks where the V2I connectivity is achieved thanks to the Road-Side Units (RSUs), as it is a prerequisite for supporting

control traffic exchange between the SDN controller and the vehicles. The presence of viable V2I wireless links is an appealing opportunity to reconsider how some core ITS services are implemented. This is, for instance, the case of the Alert Message (AM) dissemination service in an emergency or a risky situation (accident or vehicle breakdown) that we consider in this work.

In this chapter, we introduce the context of this thesis by presenting some preliminary notions, the motivations of this work, and a synthesis of data dissemination in vehicular networks. Then, we introduce our thesis scope by presenting vehicular networks' alert message dissemination problem and the general assumptions of this work. Finally, the last section concludes the chapter.

2.2 Background information

2.2.1 Cooperative Intelligent Transport System (C-ITS)

Intelligent Transport Systems (ITS) have been considered one of the main pillars of smart cities [Arroub 2016]. They are defined by The ETSI standardization organization as integrating information and communication technologies into transport infrastructures and vehicles to improve their safety, reliability, efficiency, and quality [ETSI 2020].

Cooperative Intelligent Transport System (C-ITS) refers to transport systems where the cooperation between two or more ITS subsystems (pedestrian, vehicle, and infrastructure) enables and provides an ITS service that offers better quality and an improved level of service compared to the same ITS Service provided by only one of ITS subsystems, E.g., Connected vehicles are a new paradigm of the cooperative intelligent transport system (C-ITS) aiming to improve road traffic safety and efficiency through wireless communications.

C-ITS have four main connected subsystems (Figure 2.1) [Casademont 2019]:

- Vehicular ITS station: Device equipped in vehicles called On Board Unit (OBU).
- Personal ITS station: Smartphones and other devices that aim to assist the road users (cyclists, pedestrians, ...). Frequently, these stations are called personal OBU.
- Roadside ITS station: Devices installed next to the road and called Road Side Unit (RSU).
- Central ITS station: Traffic management centers and back-office of service providers.

The cooperation of those subsystems mainly aims to reduce the number of road accidents and optimize transport time and fuel consumption to offer greener and safer transport. However, the deployment of these systems is not limited only to

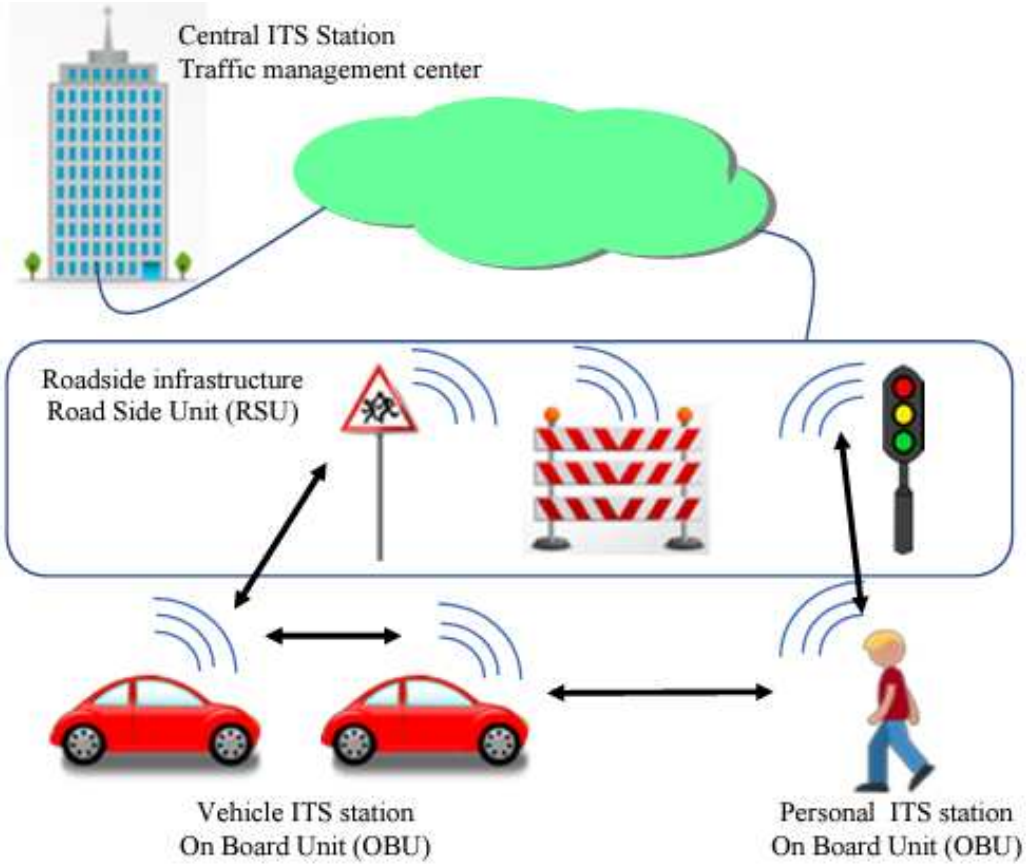


Figure 2.1: C-ITS infrastructure subsystems [Casademont 2019]

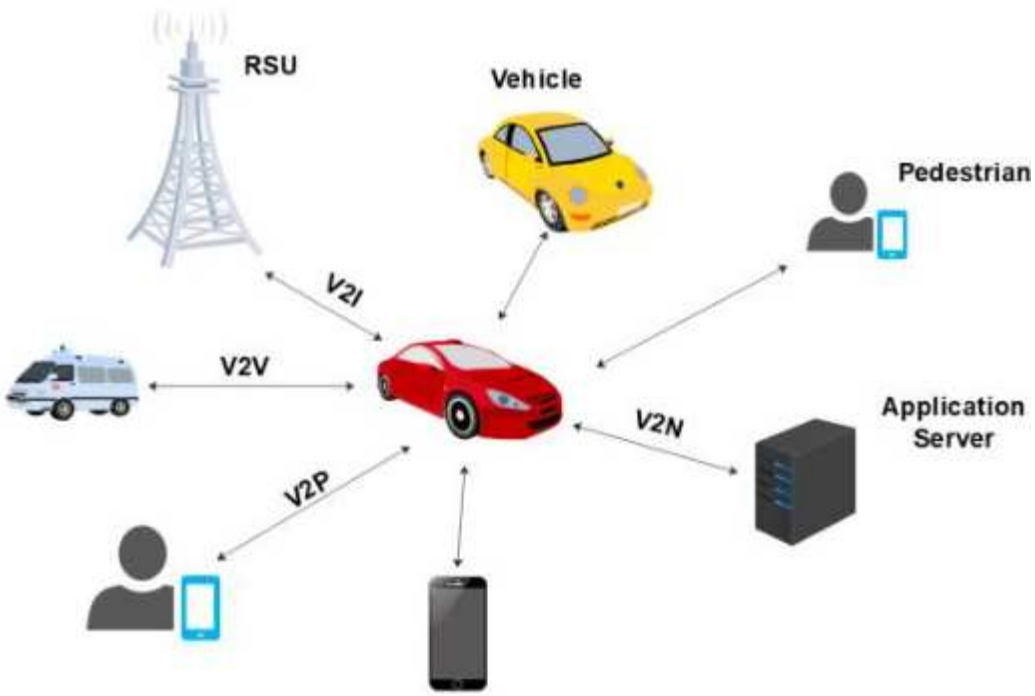


Figure 2.2: Vehicle to Everything (V2X) communications [Kawser 2019]

road transport (*considered in our studies*) but includes other areas such as rail, air, and maritime transport[ETSI 2020].

C-ITS advocates for enriching the perception that vehicles have of their surrounding world and enlarging their perception horizon are pushing toward enabling all kinds of vehicle connectivity under the umbrella of Vehicles-to-everything (V2X) communications[Sato 2020]. These latter notably include Vehicle-to-Vehicle (V2V) (*e.g., to share information about the vehicle's mobility (speed, direction, etc.)*), Vehicle-to-Infrastructure (V2I) and Infrastructure-to-Vehicle (I2V) when communication is limited to Road Side Units (RSU) (*e.g., to share information related to road conditions (warning of a danger on the road)*), and Vehicle-to-Network (V2N) communications when the vehicle communicates with an entity in the network(*e.g., a server connected to the network, or a traffic information system, cloud, etc.*). In addition, the system can also integrate exchanges between vehicles and other actors (i.e., other than the infrastructure), for example, between vehicles and pedestrians [3rd Generation Partnership Project 2015]; Vehicle-to-Pedestrian (V2P) communication occurs to signal to a pedestrian the passage of a vehicle. Figure 2.2 illustrates these various types of communications established between a vehicle and its surroundings.

2.2.2 C-ITS services

Various services and use cases have been proposed in the literature [ETSI 2009] [Network 2015]. Most of these services are defined by the ETSI standardization

organization and the 3GPP working group. We present in the following sections some examples of services. These services are generally classified into two major categories, depending on their purpose and whether or not they are related to road safety [ETSI 2009]:

2.2.2.1 Safety services

Figure 2.3 presents three services of two types of safety applications (each application includes a set of uses cases);

- *Cooperative awareness* services to warn and increase environmental awareness: two use cases: a) Emergency vehicle warning and b) Cooperative collision avoidance.

- *cooperative maneuver* services to coordinate the trajectories among vehicles: one use case: c) Platooning.

The first use case (cf. Figure 2.3, (a)) aims to warn current vehicles of an emergency vehicle (e.g., ambulance, fire department, etc.) by disseminating an alert message to all vehicles in the area. This is to inform neighboring vehicles to allow each driver to take the right actions (e.g., free the way, etc.).

The goal of the second use case (cf. Figure 2.3, (b)) is to help vehicles to avoid accidents. Each vehicle constantly transmits information about its path and state (position, speed, acceleration, etc.) and receives information from nearby vehicles. These exchanges are ensured via a V2V communication or passing through an RSU entity. The vehicles use this information to determine considered actions to avoid accidents cooperatively.

The third service (cf. Figure 2.3, (c)) allows vehicles in autonomous driving mode to provide automated overtaking while avoiding accidents. Vehicles exchange information on their trajectories and states (position, speed, etc.) in order to coordinate overtaking without accident risk in various conditions (e.g., direct road, crossroad, a road in the opposite direction).

2.2.2.2 Non-safety services

Figure 2.4 illustrates three uses cases of three non-safety applications;

- *Cooperative sensing* services to exchange sensor data and object information that increases vehicles' environmental perception: a) Bird's-eye view.

- *Cooperative traffic efficiency* services to update of routes and map: b) Itinerary recommendation.

- *Cooperative local services* to advertise and provide on-demand information to passing vehicles: c) Passenger entertainment.

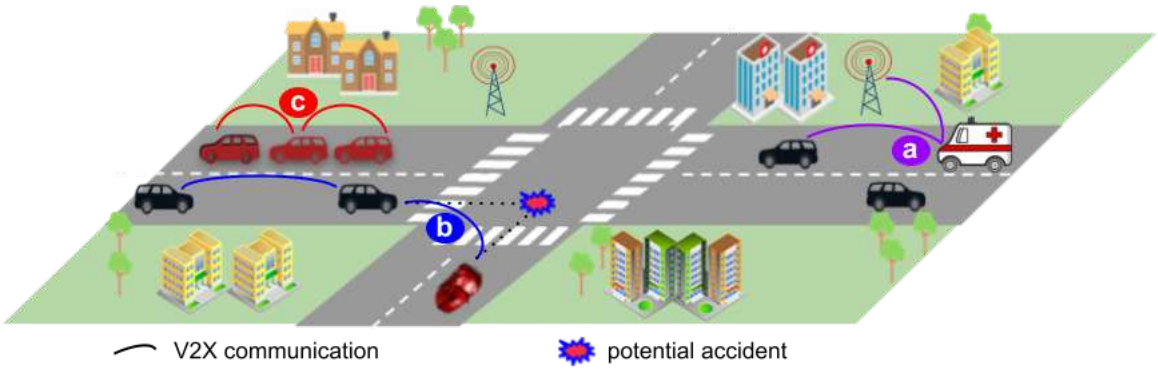


Figure 2.3: Safety use cases

In the first service (cf. Figure 2.4, (a)), vehicles share their local visions (constructed via information from onboard sensors) with neighboring vehicles; e.g., vehicles equipped with cameras can share their images with their neighbors. This allows vehicles to build a global vision of the environment (bird's-eye view) and effectively plan their future trajectories. For example, as shown in figure 2.4, the blue vehicle detects, thanks to the vision shared by the neighboring entities (RSU, vehicles), pedestrians on the road and traffic jams on its initially planned path and therefore decides to change its trajectory.

The second service (cf. Figure 2.4, (b)) makes it possible to recommend to vehicles the road to take according to a given criterion of optimality. For example, we cite the network connectivity criterion for choosing the correct route, called the most connected road [Wegner 2018]. For example, the red vehicle will receive information on the path to follow (red path), offering the best quality of network service (e.g., less overloaded network in this area).

The third use case (cf. Figure 2.4, (c)) proposes an infotainment service to improve driver and passenger comfort, where vehicles can download entertainment information via V2I/V2N communications. This multimedia content can be available locally or downloaded from a server on the Internet.

2.2.3 C-ITS service requirements

The standardization organization ETSI based on experts' recommendations and simulation results, has already defined and described key functional and performance requirements to ensure the proper functioning of C-ITS services[ETSI 2009]. These requirements are service-dependent and are expressed using the following metrics:

- End-to-End Latency (ms): Maximum delay between the moment the source application generates a data packet and the destination application receives

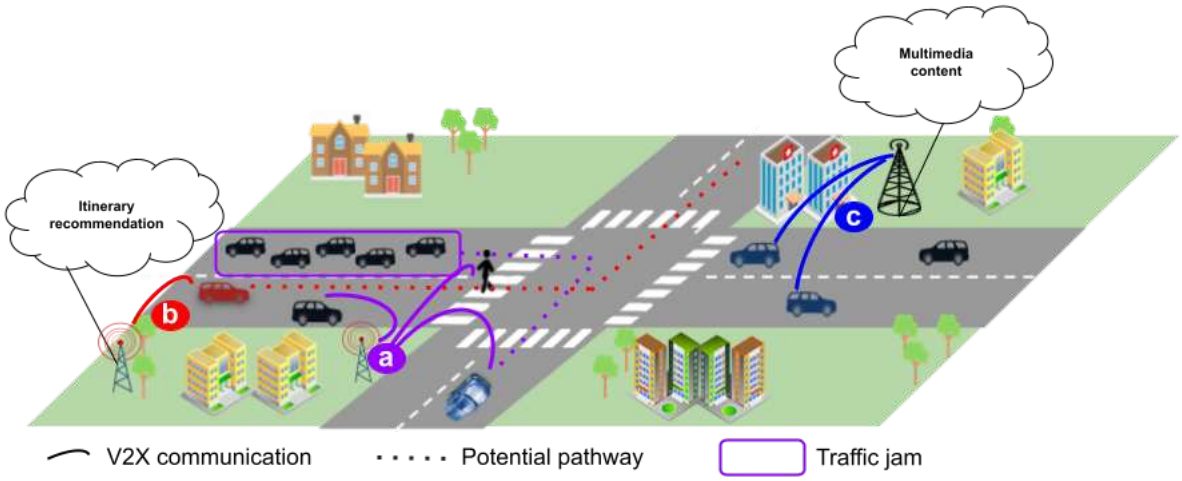


Figure 2.4: Non-Safety use cases

it.

- Reliability (%): Maximum tolerable packet loss rate measured at the application layer. A packet not received by the destination application within the maximum wait time allowed is considered lost. It is expressed in % to indicate the minimum packet success rate required (e.g., 90% means the number of received packets per the number of sent packets must be greater than or equal to 90%).
- Periodic message frequency (Hz): A sending frequency is required and expressed in Hz for services requiring periodic messages.
- Throughput ($Mbit/s$): Minimum throughput required by the application to work correctly.
- Security: Some services require specific security measurements. These include user authentication, data integrity, confidentiality, and user privacy.

Safety services are based on periodic (e.g., Cooperative Awareness Message (CAM) [ETSI 2014]) or event-driven (e.g., Decentralized Environmental Notification Message (DENM)[ETSI 2013]) broadcast messages with a repetition rate as high as 10 Hz (e.g., emergency vehicle warning) or lower (e.g., road works warning). For example, cooperative awareness services presented above require messages to be delivered with a maximum latency of 100 ms. The payload of each message ranges from 60 to 1500 Bytes [ETSI 2009], affected by the road and traffic conditions. Due to the dynamic nature of cooperative awareness applications, the data rate is between 5 kbps to 40 kbps, and the required reliability is 90-95%.

However, non-safety services are generally less strict than safety services. For example, traffic efficiency use cases are supported using the V2N or the V2I modes

Application	**V2X Mode**	**Latency**	**Reliability**	**Throughput (Kbps)**
Cooperative Awareness	V2V/V2I	$100ms$	$90-95\%$	5-96
Cooperative Maneuver	V2V/V2I	$< 3ms - 100ms$	$> 99\%$	10- 5000
Cooperative Sensing	V2V/V2I	$3ms - 1sec$	$> 95\%$	5-25000
Cooperative Traffic Efficiency	V2N/V2I	$> 1sec$	$< 90\%$	10-2000
Cooperative local services	V2N	$500ms$	$95-99\%$	> 25000

Table 2.1: Performance requirements of different C-ITS services derived from [Festag 2014]

without strict delay or reliability requirements since there is no need for prompt (re-)action at the vehicle side. Each vehicle updates the Traffic Management server (uplink) every few seconds with location, status, and road information, which are required for more efficient route selection (the payload of this type of message is 1500 Bytes [ETSI 2009]). The response from the Traffic Management servers (downlink) includes digital map updates (2 MBytes).

Table 2.1 presents the main applications of C-ITS use cases and their key performance requirements in terms of reliability, latency, and the expected data rate per vehicle.

The cooperative awareness safety services are mainly considered in our studies to illustrate the contribution of the proposed approaches.

To conclude, vehicles' communication capability and road infrastructure have opened up a wide field for the imagination of new services and use cases for ITS systems. They involve various communications; some only use V2V type communications, and others combine several types. Their requirements in terms of QoS vary depending on the kind of service and aim sought.

All C-ITS entities and their interactions form communication networks called vehicular networks, which are essential for the proper functioning of the C-ITS system. Those networks are based on norms and standards defined by international standardization organizations. Their purpose is to provide the required network connectivity with the level of performance required by the various C-ITS services.

2.2.4 Norms and standards

Two main standards are considered to support the communication needs of ITS services. Direct communication technologies are known as DSRC (802.11p in the United States and ITS-G5 in Europe) based VANET and cellular technologies (e.g., Long Term Evolution (LTE)/5G).

2.2.4.1 Dedicated Short Range Communication (DSRC) and 802.11p

The IEEE 802.11p standard is an adapted version of 802.11 wireless networks for vehicular communications. It integrates, with some modifications, the physical layer of the IEEE 802.11a standard, based on Orthogonal Frequency Division Multiplexing (OFDM), and the MAC layer of the IEEE 802.11e standard based on the Enhanced Distributed Channel Access (EDCA). It relies on the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). EDCA provides priority-based differentiated access to the transmission channel. It is enforced by controlling CSMA/CA parameters, namely the Arbitration Inter-frame Space (AIFS), which sets the free channel listening period, as well as the size of the min (resp. max) contention window size (resp. CW_{min} and CW_{max}) from which the random back-off times are derived. Compared to traditional Wi-Fi networks, one of the significant changes is removing the BSS (Basic Service Set) association step, which allows vehicles to transmit data without prior exchange of control information immediately. As a result, communication can be done in a distributed manner without dedicated network coverage. However, RSU (Road Side Unit) entities can be deployed to increase network coverage, especially in situations where direct visibility is obstructed (e.g., intersections in an urban environment), and to benefit from other network management and security services (e.g., certificate management), as well as gateways to internet access.

The DSRC standard uses multiple channels to support safety applications and other intelligent transport services. It dedicates a single Control CHannel (CCH) for safety services. In addition, six Service CHannel (SCHs) handle non-safety/infotainment service–related transmissions (e.g., video calling). Each channel exhibits 10 MHz of bandwidth within 5.9 GHz of spectrum, supporting data rates of up to 27 Mbps. Using multiple channels for data delivery is defined under the multichannel medium access control (MAC) operation of the IEEE 1609.4 standard in Wireless Access in Vehicular Environment (WAVE)[Rasool 2017].

The main advantage of DSRC is the very low latency time, under 100 milliseconds, which is ideal for safety applications that require a delay in this range. Although this technology is desirable to support V2V communications, it has limitations. Works in [Hassan 2011] [Ma 2009] show that performance drops drastically under high-density conditions. One of the main reasons for these degradations is the CSMA/CA collision avoidance mechanism. This mechanism introduces a significant delay in channel access in dense environments based on the principle of listening before transmitting (listen before talk). In addition, it suffers from the collision of hidden nodes (hidden nodes).
On the other hand, the narrow communication range of RSU entities limits their deployment for total coverage of a given area. The ITS-G5 standard is the European variant of the 802.11p standard specified by ETSI [ETSI 2019].

2.2.4.2 Cellular communications (C-V2X)

The 3GPP working group has recently been interested in developing vehicular communications, which are the core of the 3GPP Release 14 standard. Two main directions are studied. The first one concerns the LTE technology in its initial version [3rd Generation Partnership Project 2015] (already implemented by various mobile operators) to support ITS services. The second one mainly concerns the extension of this technology for direct V2V communication support (work on release 14), commonly known as LTE-V or LTE-V2X.

LTE - infrastructure based communication This involves using the infrastructure of the LTE mobile network to support vehicular communications. Each vehicle represents a User Equipment (UE) that communicates with the base station (BS) to join another entity in the network, for example, a road manager server (i.e., to retrieve information on road traffic) or another vehicle connected to the network.

Among the significant advantages of this technology for the support of ITS services, we mention:

- A large communication range of the BS entities reduces the frequency of handovers compared to RSU entities based on 802.11p.
- A high capacity network can support services with high bandwidth requirements.
- Mature technology with multiple mechanisms, for example, Multimedia Broadcast and Multicast Services (MBMS) broadcast and multicast functions that many security messaging services can use, e.g., services aimed at warning a danger on the road (i.e., accident).

However, this technology has some limitations, mainly due to the design choices initially proposed to support mobile traffic (MBB) and not vehicular traffic. Among these limitations: i) A unique radio interface connects the UE to the BS. Indeed, any data transmission (even between two vehicles close to each other) involves crossing the infrastructure, following an uplink transmission (UL) followed by a downlink transmission (DL). This limits its applicability to V2V communications, especially for safety-type services with strict latency requirements. ii) The unavailability of communications outside network coverage. Indeed, several areas are not considered initially during the deployment of LTE networks and need to be covered to support vehicular communications, for example, rural areas, mountains, etc.

LTE-V2X/ C-V2X To improve LTE technology and enhance its ability to support V2X communications effectively, release 14 (LTE-V2X) focuses on supporting direct V2V communications. In addition, a new radio interface has been specified. The PC5 interface with which two UEs (vehicles) can communicate

without resorting to the infrastructure. The standard defines two modes: V2V communication assisted or not assisted by the infrastructure.

In the first mode, the base station coordinates the communication between the vehicles. It provides vehicles with control information, specifying the resources and radio parameters to be used for their direct communications. It can also handle traffic prioritization if multiple applications are running simultaneously by the same vehicle. In this mode, signaling (control) messages are transmitted using the spectrum of mobile network operators, while data between vehicles is transmitted in a dedicated frequency spectrum in the 5.9 GHz band.

In the second mode, vehicles can configure their communications without assistance from the cellular network. They select radio resources autonomously, whether or not they are under network coverage. This mode is necessary when the vehicles are out of network coverage. They use pre-configured settings. In addition, they implement distributed algorithms for network management (scheduling, interference, etc.).

New technology improvements are being studied as part of the 5G V2X. They offer new services and use cases with more stringent requirements. The term C-V2X for Cellular-V2X has been proposed to refer to all current work and future improvements within the 5G specifications.

2.2.5 Vehicular networks

2.2.5.1 Vehicular Ad-hoc NETwork (VANET)

The main objective of VANET is to build coordination between vehicles through a distributed communication mechanism to reduce potential accidents and provide practical traffic information to passengers [Dressler 2011]. The main entities in a VANET are vehicle nodes and Road Side Units (RSUs), and they are further bridged by Vehicle-to-Vehicle (V2V), Vehicle-to-Infrastructure (V2I), Infrastructure-to-Vehicle (I2V), and Infrastructure-to-Infrastructure (I2I) communications. The theoretical communication range is 1000 meters, but in reality, it is around 400 meters for RSU, and 200 meters for vehicle [Bishop 2005]. Vehicle nodes are equipped with GPS devices and Dedicated Short Range Communication (DSRC) radios (described above). Recently, many studies have been on integrating multiple wireless technologies such as 5th Generation (5G), Long-Term Evolution (LTE), WiMax, Wi-Fi, ZigBee, etc., in the On Board Unit (OBU) of vehicles[Dressler 2011, Ku 2014, Ge 2017]. In addition, VANET employs a periodical broadcasting of Basic Safety Messages (BSMs - also referred to as hello messages) primarily used for the safety operations such as collision avoidance. VANET further holds the luxury of unique stationary RSUs in the system, whose purpose is to disseminate important information such as road conditions, traffic light status, etc., to passing vehicles and act as a relay station [Ali 2016]. The base stations connected with RSUs will provide access to the Internet. The appearance of VANETs has opened a challenge for searching for new reliable dissemination protocols and schemes with real-time

constraints intended to serve several applications related to public security. The goal is to design a protocol to broadcast a message from a single source to every node within its transmission range with the highest possible reliability and minimal latency.

VANET is a subset of Mobile Ad-hoc Network (MANET). But, in contrast to MANET, VANET exhibits several unique characteristics, such as predictable mobility patterns, frequent disconnections, high topology changing rates, no power constraints, and strict delay requirements. Moreover, VANET has often been much more extensive than MANET in network size and comprises thousands of nodes.

Even though the research on VANET has been going on for nearly two decades, it is still facing a lot of challenges such as inefficiencies in information dissemination and resource utilization, security and privacy issues, unbalanced flow traffic, unreliable connectivity, inability to cope with dynamic topology changes, which collectively result in low Packet Delivery Ratios (PDRs), high end-to-end delays and packet overheads, lack of support for QoS provisioning, etc [Dressler 2011, Jerbi 2011].

2.2.5.2 Multi-RAT vehicular networks

Each network that combines at least two diverse access technologies is called a multi-RAT network.
In vehicular networks, the C-ITS system and its various components are still under development. Indeed, the network architecture and the adopted technologies are the subjects of much work and discussion. Therefore, several criteria impact this choice. Among these criteria, we cite the maturity of each technology and its performance. The 802.11p standards are the first studied for this system, and several prototypes and products are already available. However, in contrast, it suffers from some performance limitations, reducing its ability to support a wide range of ITS services. Indeed, its latency and reliability (critical performance criteria for ITS services) decrease in case of high density and mobility of vehicles [Hassan 2011, Ma 2009, Hameed Mir 2014]. Moreover, this technology is prone to interference and collisions due to hidden nodes.

On the other hand, cellular communications represent a promising technology to support ITS services, especially those requiring V2N communication. They are characterized by high network capacity and wide coverage. Performance tests of LTE technology for vehicular communication support show that it mostly meets the constraints of reliability, throughput, and network coverage [Hameed Mir 2014, Liu 2018b]. However, its performance also degrades under high density conditions [Hameed Mir 2014, Liu 2018b, ETSI].
Several works conclude on the limit of this technology to support safety applications with stringent requirements in terms of latency. This limitation is mainly because the communications must pass through the base station before arriving at their final destination (the main subject of release 14 with the proposed direct communication mode). Several improvements are under development, including support for direct

communication or V2V (from release 14) and other features as part of the 5G specifications.

The main vision shared by the community is to adopt a hybridization of the two complementary technologies, namely DSRC and cellular communications, given the complementary nature expressed by most of the performance analysis work [Hameed Mir 2014, Liu 2018b]. Hence, what is called multi-RAT diversity, combines at least two diverse access technologies. This makes it possible to take advantage of the complementary benefits of the two technologies. Therefore, it is possible to favor DSRC for V2V communications and cellular technology for V2I/V2N communications. However, this hybridization presents some challenges, notably the coexistence of these technologies since various actors should manage them and the choice of the network to be used when considering the latter's various characteristics and the ITS services' needs.

2.2.5.3 Software Defined Vehicular Network (SDVN)

Recently, researchers have focused on incorporating Software Defined Networking (SDN) into vehicular networks to address problems mentioned above, enable flexibility and programmability, and encourage innovation [Ku 2014, He 2016].

In a network, each network device comprises a data plane and a control plane. The main purpose of the data plane is to route data, while the control plane takes care of all network control decisions, such as deciding which interface to route data. In a conventional network, the data plane and control plane are integrated within the same device, and each device decides locally. However, the SDN paradigm advocates the idea of separating the data plane and the control plane. Indeed, network control functions are externalized from the network devices and arranged in software components on dedicated external devices called SDN controllers, as illustrated in figure 2.5.

In this architecture, the SDN controller is the central element. It communicates with the different nodes of the network through a SouthBound Interface (SBI) protocol - the most widespread is the OpenFlow standard - while the applications express their needs to the SDN controller using the NorthBound Interface (NBI).

The controller provides an abstraction of the underlying network to network applications and services and implements various network policies. The data plane comprises forwarding nodes, often called Packet Forwarding Elements (PFEs). They forward packets according to the rules proactively or reactively installed by the SDN controllers.

This paradigm has initially revolutionized wired network architectures and has been widely adopted in most infrastructures (e.g., data center networks, campuses, etc.). The success and achievements it has shown in these networks have attracted the community's attention, and several works are interested in its adoption in other types of networks. It is notably true for vehicular networks. These networks form the fundamental basis of an intelligent transport system. Various scientific research organizations, industries, and standardization organizations are interested

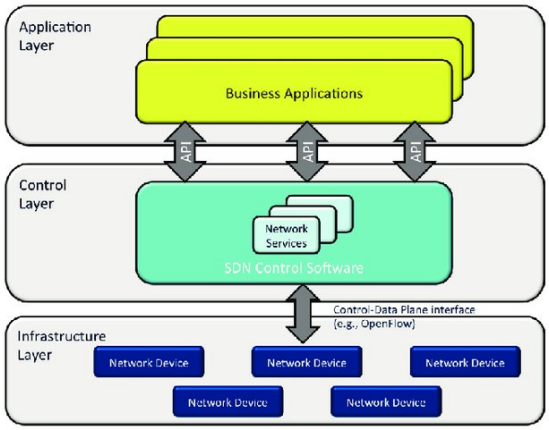


Figure 2.5: Global architecture of the SDN paradigm, ONF [Foundation 2012]

in improving them by proposing new architectures and new mechanisms in order to support ITS services effectively. SDN paves the way for effective network control applications to unify the control of heterogeneous multi-access vehicular networks: it can leverage the vehicular network's centralized, multi-level, and multi-access view with complementary vehicles, roads, and routes traffic-related information brought from the cloud. SDN can handle the dynamic nature of networks with intelligent applications while lowering operational costs through simplified hardware and software. Moreover, SDN is a key paradigm for supporting the diverse QoS expectations (some very stringent) of C-ITS services.

2.2.6 Data dissemination strategies in vehicular networks

C-ITS services improve drivers' and passengers' security and comfort on roads by notifying dangerous situations and entertainment (e.g., parking, hotel, etc.), sharing multimedia content, etc. Generally, these services are based on periodic data dissemination as point-to-multi-point communications and geo-Broadcasting. This is to enable the state of the road and surrounding vehicles.

An effective dissemination solution for Vehicular networks must consider the latter's characteristics, such as the size of the network, the speed of vehicles, the intermittent connection of the network, and the different needs of the applications in terms of QoS, whatever the service type (safety or non-safety) and the data to be disseminated.

All data dissemination schemes should satisfy all the following objectives [ETSI 2009]: (1) High reliability: high reception rate; (2) Low latency: the duration from the first transmission attempt to the end of the broadcast step, should be as low as possible; (3) Low probability of collision: the protocol should suffer from collisions to the minimum possible; (4) Hidden node problem: avoiding the hidden node caused by the receiving collisions.

In the literature, several strategies have been proposed. Each of them may require one or more hops for routing its data, as well as the deployment or not of

infrastructure, such as roadside units (RSUs). However, all strategies rely on the cooperation of network vehicles to relay messages. Below, we will detail a general classification of dissemination strategies encountered.

2.2.6.1 Flooding

One of the most used approaches for disseminating data in VANETs is flooding. It can be used with a single hop and several hops. A message sent by a transmitting vehicle by broadcasting is transmitted to all its direct neighbors, then retransmitted once again by each receiver until reaching all intended receivers. This approach does not require any prior information about the neighbors of the vehicle, which allows it to ignore the non-existence or inaccuracy of information about the network topology. Furthermore, it increases the delivery rate and improves data transmission speed because a receiver vehicle receives several copies of the message, arriving through several routes. However, this approach also increases competition for access to the communication channel and bandwidth utilization, which does not allow it to be scaled up at the risk of generating intense network congestion [Tseng 2002]. Citing some examples:

The study's authors [Korkmaz 2004] propose a multi-hop broadcast protocol for urban environments to remedy the problems associated with the massive broadcast. A transmitting vehicle broadcasts the message to its direct neighbors; only the furthest vehicle rebroadcasts it. When meeting an intersection, vehicles are selected as relays and rebroadcast information on the different segments of the intersection.

The authors of the solution [Sun 2000] adopt the same dissemination approach. First, a message is broadcasted to reach a certain group of vehicles. However, from the second transmission of the message, only the vehicles on the edges are selected as re-broadcasters.

The criteria for selecting re-broadcasters in these two approaches mainly concern their geographical positions from the sender. However, this does not respond to VANETs issues, such as adapting to the changing density of the network, because no relationship between the number of relays and the density is given.

2.2.6.2 Probabilistic

This approach attempts to reduce the redundant messages generated by calculating the probabilities of encounters between two vehicles before deciding on the information dissemination path without requiring knowledge of the network topology. Instead, a vehicle using this approach can rely on its knowledge of the network, its history of encounters with other vehicles, and the information it has collected on the mobility and locations of other vehicles in the network.

For example, the solution [Palazzi 2012] uses this probabilistic approach; the decisions concerning the choice of relay vehicles for the retransmission of a message are based on the probabilities of encounters with the recipient vehicle(s). In solutions [Wisitpongphan 2007] and [Slavik 2010], the receiving vehicles of a

message calculate their retransmission probability based on the distance separating them from the transmitting vehicle. The greater the distance is, the greater the probability of retransmission. The authors of the solution [Busanelli 2009] use the criterion of the distance between a receiver vehicle and a source vehicle to calculate the probability of retransmission and add a parameter concerning the local density of the network, i.e., the number of direct neighbors of the receiver vehicle, to reduce the number of relay vehicles when the density is high.

2.2.6.3 Geographic

This dissemination approach is based on geographic routing protocols, where the location information of the vehicles in the control messages is periodically broadcasted in the network when it follows a pro-active approach [Bakhouya 2009] or broadcasted on demand when it follows a reactive approach [Lee 2010]. Each vehicle regularly updates a table containing its neighbors' location history to route its messages by the shortest path and, thus, reduce their routing delay. For this reason, during dissemination, the vehicle closest to the receiver(s) is selected at each hop. This approach also makes it possible to target a group of vehicles thanks to their geographical coordinates, as is done in applications designed to warn drivers of the risk of collision at intersections, for example.

2.2.6.4 Channel resource oriented

As the resources of the communication channel are limited, accessing the channel and allocating its resources becomes an optimization problem. However, this problem is likely NP-complete, considering all the variants that must be considered and the limited network information available to the vehicle.

Some solutions propose algorithms based on heuristics, such as the study [Palma 2012], which proposes a routing of access data by considering the history of encounters of the transmitting vehicle with other vehicles in the network. The aim is to estimate potential congestion and network density and then consider them to improve the delivery rate and limit the number of duplicate messages. In the solution [Liu 2007], each node maintains a table with information about the throughput and channel conditions in order to choose which relaying node is preferable to transmit its message through. However, these solutions require message exchanges between vehicles to maintain control over the use of channel resources.

Another solution [Peng 2007] improves the reception rate of emergency messages by allocating a part of the available bandwidth to them. In this solution, each node first sends a signal as a pulse, then transmits its emergency message.

2.2.6.5 Message priority oriented

In order to meet the different QoS needs for the multiple applications of VANETs, dissemination solutions propose an adaptation of the dissemination with the

importance of the content of the messages exchanged. For example, do not systematically delete all new incoming messages in case of network congestion. The solution [Torrent-Moreno 2004] solves this problem by setting priorities for access to the communication channel according to the access categories ACs, set by EDCA [Committee 1999], for each message. Another solution [Gerharz 2003] allocates tokens to the queues formed by messages wishing to access the channel. It manages channel access by weighing the number of tokens offered against the channel density and message priority. Like the latter, the solution [Rahim 2009] schedules the messages to be sent based on the channel's available resources and the message's importance, using a system of waiting queues where a higher priority is given to the most urgent messages.

2.2.7 Synthesis

VANETs are constrained by memory space, computing capacity, and energy issues. However, they suffer from a large amount of data to be sent and the large geographical areas to be covered. These, combined with the dispersion and high mobility of vehicles, the lack or insufficient infrastructure, and the variable density of the network, create several problems for disseminating data. Therefore, they cannot meet the high demands of C-ITS services, especially safety services such as cooperative awareness services, which require, in most cases, effective and reliable information dissemination. In order to address these problems, researchers have recently focused on incorporating SDN into VANET to enable and facilitate hybridization with cellular networks. As mentioned above, this incorporation allows for managing these networks and optimizes their performance.

We consider such an SDN-enabled multi-access vehicular network in our work to propose an efficient dissemination scheme for critical safety services. Under this assumption, Vehicle to Infrastructure (V2I) wireless connectivity is nominally available, with infrequent and time-limited connectivity losses. This holds for LTE and DSRC-based access networks where the V2I connectivity is achieved thanks to the RSUs, as it is a prerequisite for the support of control traffic exchange between the SDN controller and the vehicles.

The presence of viable V2I wireless links is an appealing opportunity to reconsider how some core ITS services are implemented. This is, for instance, the case of the Alert Message (AM) dissemination service in an emergency or a risky situation (accident or vehicle breakdown) that we consider throughout our work. This service will be detailed in the following section.

2.3 Thesis scope

The objective of this thesis work is to define a new solution for alert message dissemination in Software Defined Vehicular networks (SDVN) architecture with QoS, such as reliability, delay, and information coverage, as well as the analysis of the performance of this solution.

This section introduces the problem of alert message dissemination, an overview of the proposed solution, and general assumptions.

2.3.1 Alert message dissemination in vehicular networks

The Alert Messages (AM) are used to urgently warn vehicles heading into a geographic area that hosts a hazardous situation or event (an accident, tricky weather/road condition, sudden break, etc. AM service broadcasts on an area of interest alarm messages for some period to let close-by vehicles react quickly and adequately to the risky situation [Salim 2007].

AMs belong to the Decentralized Environmental Notifications Message (DENM) class of messages defined to convey the information related to different road events that need to be delivered to road safety applications of nearby vehicles [ETSI 2013]. They usually include the type of event, the coordinates of the event's position, an indication of the geographic area of the distributed message localization, and the sender. Their sizes are short (between 200 to 1200 bytes) to ensure the rapidity of their transmission. A road event is typically detected by one or multiple vehicles; multiple DENM messages related to the same event can be sourced from different vehicles. Also, an ITS application running behind the road infrastructure can detect the event and send DENMs to vehicles. When a road event is detected, the corresponding DENM is periodically generated in an average of 10 messages per second for the period where the event remains present. The dissemination requirements of a DENM message are typically expressed in terms of a geographic area of interest, which delimits the area within which all the vehicles should receive the message, and delivery/dissemination delays [ETSI 2009]. For instance, for AMs, the area of interest is a few kilometers from the event's location, meaning that multiple wireless transmission hops could be needed to cover the entire area, while the dissemination delay requirement is a few hundred milliseconds.
Such reliability and delay expectations heavily depend on the performance of the wireless technology. Therefore, the DSRC protocol is the leading vehicular communication technology recommended for supporting this kind of service [ETSI 2019] [Kenney 2011]. As explained above, it operates on a dedicated 5.9 GHz unlicensed band, split into one control channel (CCH) and 6 (or 5) Service channels (SCH). The CCH is used for DENM transmissions with another critical class of Cooperative Awareness Message (CAM) messages. CAMs are periodically broadcasted by each vehicle to inform all their wireless one-hop away neighbors (vehicles and RSUs) of their current status, e.g., position, speed, heading, etc.

Usually, AMs are relayed over multiple wireless hops to reach all vehicles within the geographic area of interest. However, rebroadcasting AMs leads to increased medium contention, transmission collisions among neighboring nodes, and useless duplicate transmissions because of the shared wireless medium. This leads to inefficient use of radio resources and increased dissemination delays. High vehicle density exacerbates these problems due to excessive redundant and contending rebroadcasts (named in the literature as the broadcast storm problem). As a result,

AMs delivery can be compromised for some time, disrupting the service provided by road safety applications.

In our work, we propose a novel Location-Alert Message Dissemination scheme for future vehicular networks that exploit V2I links to widely broadcast alert messages complemented with parsimonious V2V rebroadcasts. The centralized network controller drives V2V rebroadcast selection to reach the vehicles suffering from a potential loss of V2I connectivity while drastically controlling and reducing the number of rebroadcasts. It relies on the centralized view that the SDN controller builds on its coverage area and information related to road traffic and vehicles (their characteristics and potentially their road trip).

2.3.2 Assumptions

This section presents the assumptions and key principles of the considered architecture in this work.

The network controller represents the central part of the architecture. It holds the network's intelligence, manages the RSU/BS network, and implements the various network policies. A network controller can be an SDN controllers as described in [Toufga 2017], a centralized location server [Katsaros 2013], a fog/cloud computing [Ning 2019], or a simple central RSU/BS controller.

In our work, we have considered the architecture proposed in [Toufga 2017], whose data plane comprises vehicles, RSU entities, and base stations, all programmable via SDN. The SDN controllers represent the central part of the architecture. They host all the network control functions to define the various rules communicated to the data plane nodes. They consider three main controllers: the first to manage the cellular network, another to manage the RSU network, and the last one (the main) to ensure the coordination between the different controllers. Indeed, the main second-level controller builds a global view of the communication infrastructure using the information sent by the controller of each network. It defines and sends to each controller the global rules that describe the general behavior of the network, while the local first-level controllers (BS and RSU controllers) define the specific rules to be implemented by each network node. However, some network control decisions can be made by local controllers and do not require directives from the global controller, for example, horizontal handover operations (change of attachment point within the same network, e.g., RSU<->RSU).

As mentioned above, throughout this work, we consider only safety services that require disseminating Alert Messages (AM) with very low latency and high reliability. As a result, all vehicles will communicate with the same network interface to share the message. Preferably, the DSRC interface, thanks to its very low latency for sending this kind of small-size critical safety message, as explained above. For this and other technical reasons, we have only considered the RSU network architecture (DSRC) in this work. The idea behind the presentation and choice of a hybrid and centralized architecture is to avoid neglecting other communication

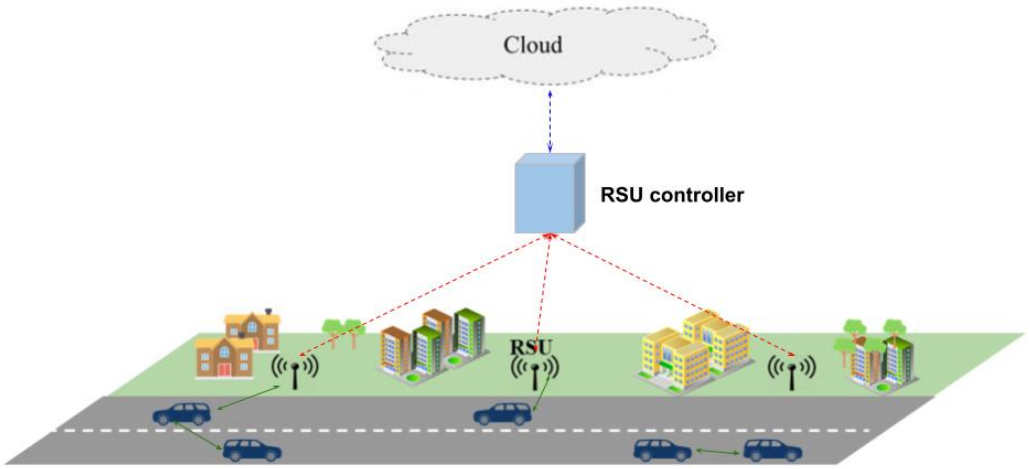


Figure 2.6: Considered architecture overview

technologies because they are complementary, and each can serve a specific type of service effectively. However, the proposed solutions can be applied to other services and networks, such as cellular networks.
Figure 2.6 illustrates the overall vision of the considered architecture.

We consider vehicles can be equipped with one or more network interfaces. For example, one interface is to access the RSU network (DSRC) (by default) for the exchange of CAMs and DENMs messages, and another one is to access the cellular network (LTE/ 5G) for the exchange of multimedia content, itinerancy, etc.

2.4 Conclusion

This chapter has introduced an overview of the background landscape necessary for understanding the thesis scope. After, we studied the techniques used for general data dissemination to ensure more or less the QoS required. Further, we have explicitly presented the alert message dissemination in vehicular networks and their strong requirements.

In this context, it is fundamental to introduce the different vehicular communication technologies to address the above-mentioned limitations of disseminating these critical messages. To this end, disseminating alert messages must offer highly controlled (ideally full) information coverage while reducing the number of rebroadcasts. In addition, the dissemination scheme should drastically reduce collisions and significantly save network resources, as useless rebroadcasts are avoided. Last, this scheme needs to induce insignificant signaling and network overhead.

In the following chapter, we present our contributions to these primary challenges.

Chapter 3

LAMD: Location-based Alert Message Dissemination Scheme

3.1 Introduction

Point-to-multi-point communications and, more specifically, Geo-Broadcasting are at the heart of many C-ITS services. Alert Message (AM) dissemination is such a service which, upon an emergency or a risky situation (accident or vehicle breakdown), alert messages are generated for some time and spread over a geographical area of interest as fast as possible to allow other vehicles to react appropriately very quickly [Salim 2007]. Historically, AM dissemination was carried out by a DSRC-based VANET operating on a dedicated communication channel, inviting vehicles to rebroadcast the alerts blindly. This method, known as simple flooding, causes a huge amount of unnecessary re-transmissions and collisions (known as the broadcast storm problem), thus wasting bandwidth, increasing dissemination delay, and lowering packet delivery ratio [Tseng 2002]. Many proposals from the literature tried to mitigate the broadcast storm problem by controlling the vehicles allowed to rebroadcast (using some form of random selection or clustering techniques to group vehicles) or reducing the contention. But, most were designed for VANETs and neither assume the presence of V2I links nor take advantage of the benefits that a global view of the network at a centralized controller can bring to the dissemination scheme.

The fundamental objective of this chapter is to develop a novel Location-Alert Message Dissemination scheme called LAMD for future vehicular networks that primarily exploits V2I links to widely broadcast alert messages complemented with parsimonious V2V rebroadcasts in order to reach vehicles located in areas that are poorly or not covered (gray zones) by transmissions from the infrastructure. A centralized network controller drives V2V rebroadcasts selection to reach the vehicles suffering from a potential loss of V2I connectivity while drastically controlling and reducing the number of rebroadcasts. It relies on the centralized view that the network controller builds on its coverage area and control as well as information related to road traffic and vehicles (their characteristics and potentially their road trip). More precisely, the network controller computes rebroadcast locations (or rebroadcast points) to ensure that V2I broadcasts and V2V rebroadcasts lead to the full coverage of an area of interest. The selection process of vehicles allowed to rebroadcast an AM is finally based on their close vicinity to rebroadcast points. This alert message dissemination scheme offers highly

controlled (ideally full) information coverage while drastically reducing the number of rebroadcasts compared to existing techniques. This considerably reduces the number of collisions and significantly saves network resources; useless rebroadcasts are avoided. Finally, our proposal induces insignificant network signaling overhead.

The rest of the chapter is organized as follows: Section 3.2 presents an overview of existing work that aims at improving AM dissemination. Section 3.3 describes the proposed location-based dissemination algorithm and the system model. Section 3.4 evaluates and discusses our simulation results. Section 3.5 concludes the paper.

3.2 Related work

Alert message dissemination has been widely studied in a VANET/V2V context as referenced by survey papers [Ullah 2018, Sanguesa 2016, Ibrahim 2020, Shahwani 2021]. All aim to address the "broadcast storm" problem and ensure high information coverage. The way they handle the problem can be broadly classified as follows. Either by controlling the flooding procedure or preventing systematic blind rebroadcasts from each receiving vehicle. Or by organizing vehicles in clusters and structuring the alarm message dissemination along these clusters avoiding the broadcast logic. Or finally, by using routing protocols to compute point-to-multi-point routes along which alarm messages are forwarded.
Below, we dive further into the details of our classification and present, without being exhaustive, some representative papers that fall in each class.

3.2.1 Control of the flooding procedure

A lot of the proposed schemes try to control the flooding procedure of AMs either by reducing the number of rebroadcasters or by reducing the contention between transmitting vehicles. They are described hereafter.

3.2.1.1 Relay selection

In these methods, only part of the vehicles is selected to relay the AMs. This selection is typically based on a combination of vehicles' characteristics, distance from the sender, local density, interests (e.g., the destination district), transmission power, etc., and helps to limit re-transmissions and contention at the cost of reduced information coverage.

Adaptive Data Dissemination Protocol (AddP) [Oliveira 2017] selects forwarding nodes based on the local density of the forwarder and the positions of neighboring nodes. Since forwarding nodes are chosen during the dissemination of each multicast packet, the nodes closest to the boundaries of the communication range and the highest local density have priority for broadcasting the message. AddP helps mitigate the broadcast storm problem (by reducing redundant messages and the hidden node problem) and frequent disconnections. However, the proposed mechanism may not be suitable for highly mobile and dense V2V networks in which

local density and neighborhood of vehicles keep changing rapidly, and many vehicles are multicast recipients.

The idea of Liu and al. in [Liu 2015] is to combine the advantages of various communications and cloud communication technologies to address the downlink data dissemination strategy from a remote server to a targeted area by ensuring reduced delays and high packet delivery rates. They propose a framework for the dissemination of alert messages within an integrated system which comprises a Hybrid VANET-Cellular architecture where the buses act as mobile gateways (vehicles with dual DSRC and Cellular interfaces, registered in the cloud) and a cloud infrastructure which enables rapid data acquisition of road traffic flow and the geographical position of all mobile gateways. This choice efficiently provides essential traffic information (accident, route recommendation, etc.) to the vehicles in the targeted area. Gateways play the role of rebroadcasters by relaying back and forth to the farthest receivers ensuring that remote vehicles can immediately get the alert message. Since some gateways may broadcast the message simultaneously, the gateway relay selection process must be interrupted when the vehicles ahead detect that they have already received the message; thus, the data redundancy can be considerably reduced. But this method remains restrictive because it requires that the nodes be spread over the targeted coverage area.

Black-burst and Multi-channel-based Multi-hop Broadcast protocol BMMB [Wu 2017] runs on different channels simultaneously. The black burst is based on splitting the communication range of the sender of the message into several segments via reiterative partitions; then, using the Request to Broadcast and Clear to Broadcast (RTB / CTB) mechanisms, a single relay vehicle is selected in the outermost non-empty segment to broadcast the emergency message on the next hop. This protocol minimizes the unwanted effects of the broadcast storm but assumes non-standard wireless technologies.

3.2.1.2 Adjusting of MAC layer parameters

In this second sub-class of methods, MAC level protocol parameters are adjusted on a vehicle basis for statistically assigning different back-off periods to vehicles. This choice reduces node contention and collisions with no impact on information coverage. However, useless transmissions are not avoided.

In [Virdaus 2017], Virdaus and al. propose a new model to calculate the survival probability of a single-hop broadcast packet. This model uses a forwarding mechanism in an emergency message dissemination application based on counting all the cases of contention window allocations to all the nodes simultaneously receiving a broadcast message. Their scheme communicates a single-hop broadcast survival probability as a function of the size of a contention window and the number of broadcasting nodes. The proposed model improves the messages reception ratio. However, there is a significant increase in the dissemination delay.

In [Balon 2006], the authors propose a protocol that minimizes the collision rate and increases the broadcast reliability. A node can detect collisions and congestion

by simply analyzing the sequence numbers of the recently received packets. Each node periodically broadcasts its status to its neighbors, e.g., relevant information on its position, speed, acceleration, etc.) and, notably, a report on the portion of packets it successfully received from each neighbor. Accordingly, based on this feedback mechanism, each node dynamically adjusts its parameters, such as its Contention Window (CW) size and transmission rate.

In [Gawas 2017], an efficient time-slotted multi-hop broadcast protocol is proposed. This protocol selects a segment leader responsible for forwarding the alert message on a particular road segment. It allocates separate time slots for alert messages to avoid interference and ensures reliable delivery through the signaling mechanism while maintaining a high reception rate and low end-to-end delay for single-hop safety messages. This protocol also reduces the number of transmissions and ensures the timely delivery of warning messages.

3.2.2 Vehicle clustering techniques

Another class of proposals set aside the flooding logic and guide AMs dissemination by organizing vehicles into clusters (groups of nodes) and defining how messages are propagated between and within clusters. Authors use many specific criteria for cluster formation and Cluster-Head (CH) selection as neighborhood, direction/destination, relative velocity, etc.

In [Liu 2018a], clusters are vehicles traveling in the same direction, with similar speeds (for cluster stability) and mutual radio coverage. The node with the highest link quality in the cluster is elected as CH, receiving messages from other CHs and retransmitting them to other cluster members. The proposed method considerably improves the emergency message delivery ratio. However, there is a significant increase in the delivery time.

In [Ali 2019], they define clusters as a set of vehicles around a nearby CH vehicle (i.e., with mutual radio coverage with CH), having the same direction, the same destination, and the same interests. Each CH maintains a list of neighboring CHs, including clusters from other roads. When a CH receives an emergency message with the same road ID as its own, it sends it to all its cluster members and all neighboring CHs (within its transmission range R) moving in the opposite direction or on a different road ID. The information is exchanged through beacons with no beacon congestion control mechanism.

Cheng and al. [Cheng 2019] have proposed a new Center-Based Stable Clustering algorithm (CBSC). They assume all vehicles have LTE and DSRC interfaces for communication. The main idea of their proposal is to detect the center of the ranges where the density of vehicles is much higher than in other areas with the help of receiving beacon messages. For stable cluster formation, the vehicle grouping is based on these detected center ranges with the help of the blob detection method [Lindeberg 1998]. An improved highest-degree algorithm is applied where the density of vehicles is low. A new relative mobility metric is introduced for the CH election that evaluates the speed, position, and maximal

acceleration differences between one vehicle and all remaining vehicles of the same cluster. A vehicle with a smaller value of relative mobility metric works as CH. The LTE-enabled infrastructure unit (eNodeB) manages many of the formed clusters due to the high coverage radius of LTE.

3.2.3 Routing protocols

The third class is based on routing protocols that proactively or re-actively compute point-to-multi-point routes to all known vehicles. As a result, they exhibit the most predictable information coverage and efficient resource usage, but their use creates a higher network overhead and an increased complexity embedded in vehicles.

The ZoomOut Broadcast Routing Protocol (ZBRP) broadcast routing protocol developed in [Aldabbagh 2016] provides security details to the VANET driver with a rear sensor. It detects messages for the neighbors with a simple procedure: to use one-hop intelligently depending on vehicles' speed and distance to define a front and a rear vehicle. As multicast security message delivery occurs, an entry or relay is received when unrelated parties lose a safety note. During processing multi-hop safety messages, only a front or a rear relative vehicle rebroadcasts a safety message, whereas non-relatives drop it.

The Ad hoc On-Demand Distance Vector (AODV) routing protocol [Narang 2015] supports on-demand, unicast, and multicast routing, using a sequence number for each route. AODV responds to link failure in the network because it is responsive. However, its major drawback is that this type of network can determine a route that is not necessarily optimal. Therefore, an enhancement to AODV, Ad hoc On-demand Multipath Distance Vector Routing (SD-AOMDV), is proposed in [Maowad 2012] to address the characteristics of VANETs. SD-AOMDV adds additional criteria (speed and direction) to the field that determines the next hop during finding and building routes; when a source node wants to send a packet to destination nodes, the routing protocol gets the direction and speed of the source node. Then, it gets the direction and speed of the destination nodes. Finally, intermediate nodes taking part in the route between source and destination are specified based on the direction and speed of both source and destination.

Preferred Group Broadcasting (PGB) [Naumov 2006] is a greedy forwarding technique that significantly improves geographic routing protocols to prevent broadcast storms and hidden node problems while disseminating messages. PGB addresses these problems by allowing only specific nodes to rebroadcast a message. The PGB algorithm selects relay nodes and classifies each node that receives a packet into one of the three groups: PG (Preferred) group that is composed of pre-defined nodes, such as ambulance, police vehicles, etc.; IN group composed of nodes with a signal stronger than PG nodes; and OUT group composed of nodes with a signal weaker than PG nodes. Nodes from PG have the highest priority to be chosen as relays, nodes from the OUT group, and finally from the IN group. PGB uses signal strength to calculate waiting-timeouts. The highest priority relays will set shorter waiting- timeouts. One of problems of this protocol is that it lacks

a mechanism to avoid message duplications. Another disadvantage is that, in low-density scenarios, the message rebroadcasting can be discontinued in case of an empty vehicle group (network partition).

Our proposed method, LAMD, is a relay selection technique that differs significantly from other previous works because it considers only one parameter: the location of vehicles, making our method more efficient and simple. In addition, and as mentioned above, our dissemination scheme primarily exploits V2I transmissions and complements V2V re-transmissions to reach vehicles in gray areas, while most of the other work is solely concerned with V2V communications. Table 3.1 presents a comparison of reliable protocols, considering the following characteristics: (1) communication support; (2) class of method used to disseminate the message; (3) parameters considered to select relay and cluster head/member or Adjust Mac parameters (4) the primary requirement metrics to disseminate an alert message: Information coverage (to ensure that all vehicles in the region have received the message), Dissemination Delay (to ensure that the message has been received by all vehicles quickly), Collision ratio (to ensure that the broadcast storm problem has been mitigated and contention are reduced). Conversely, to all previously analyzed work, Table 3.1 clearly shows that our LAMD protocol significantly improves the information coverage, dissemination delay, and collision ratio.

Table 3.1: Comparison of related research

Selected work	**Communi -cation**	**Dissemination technique**	**Considered parameters**	**Information coverage**	**Dissemina -tion Delay**	**Collision Ratio**
ZBRP [Aldabbagh 2016]	V2V	Routing	vehicles' speed and distance	✓	✓	✓✓✓
D-AOMDV [Maowad 2012]	V2V	Routing	vehicles' speed, direction and distance	✓	✓	✓✓✓
PGP [Naumov 2006]	V2V	Routing	vehicles' type and signal strength	✓	✓	✓✓✓
Liu and al. [Liu 2018a]	V2V	Clustering	vehicles' link quality and distance	✓	✓	✓✓
Ali and al. [Ali 2019]	V2V	Clustering	vehicles' distance, direction and interests	✓	✓	✓✓
CBCS [Cheng 2019]	V2V	Clustering	vehicles' speed, position, and maximal acceleration	✓	✓	✓✓
Virdaus and al. [Virdaus 2017]	V2V	Adjusting MAC Layer parameters	Contention window	✓	✓	✓✓✓
Balon and al. [Balon 2006]	V2V	Adjusting MAC Layer parameters	Contention window and vehicles' transmission rate	✓	✓	✓✓✓
Gawas and al. [Gawas 2017]	V2V	Adjusting MAC Layer parameters	Time slot and vehicles' transmission rate	✓	✓	✓✓✓
BMMB [Wu 2017]	V2V	Relay Selection	distance, vehicles' distance and position vs sender	✓	✓	✓✓✓
Liu and al. [Liu 2015]	V2I/V2V	Relay Selection	vehicles' distance and position vs sender	✓	✓✓	✓✓✓
Addp [Oliveira 2017]	V2V	Relay selection	vehicles' distance and local density	✓✓	✓✓✓	✓✓
LAMD	V2I/V2V	Relay Selection	vehicles' Location	✓✓✓	✓✓✓	✓✓✓

(✓) : Slightly enhanced ; (✓✓) : Moderately enhanced; (✓✓✓) : Highly enhanced

3.3 Location-based Alert Messages Dissemination scheme (LAMD)

3.3.1 System model

An SDN network controller is assigned to each region; it manages all the RSUs that provide V2I wireless connectivity within its region. Each controller identifies gray zones in its region and selects K rebroadcast zones where vehicles can rebroadcast an AM sent by an RSU to extend the V2I coverage and reach all the vehicles located in gray zones. These rebroadcast zones are defined by a rebroadcast point $P_i(x_i, y_i)$, $i \in [1, K]$, where x_i and y_i are the GPS coordinates of P_i, and a radius d_{max} in the order of a few meters from the rebroadcast point.

For illustration, Figure 3.1 depicts a road section involving two RSUs led by an SDN

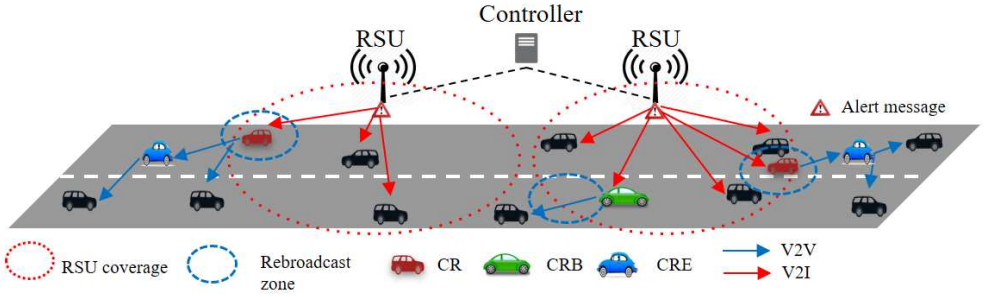


Figure 3.1: Location-based alert messages dissemination scenario

controller. The red dash ellipses represent RSUs' wireless coverage. All the other road portions are considered gray zones. Three rebroadcast zones (delimited by the blue dashed ellipses) are defined. Vehicle transmissions from these rebroadcast zones are expected to cover the gray zones.

3.3.2 LAMD key principles

LAMD is an alert message dissemination procedure designed for network-infrastructure-based vehicular networks, typically SDVN. It relies on V2I broadcasts from the network infrastructure (i.e., RSUs) complemented with V2V rebroadcasts aiming to reach and cover all vehicles in a geographical area of interest, notably those that do not fall under the coverage of the RSUs. The key principles of LAMD are as follows:

The network controller computes rebroadcast points/zones used to select vehicles allowed to undertake V2V retransmissions effectively. It shares these zones with vehicles during handover. LAMD is focused on dissemination, so it doesn't preclude any particular rebroadcast point placement method. This latter can be based on network planning, radio site surveying, or any form of prediction to cope with the dynamic nature of radio channels. At first, a general simple default method

is, however, proposed in the next section. Later, we offer a more optimal rebroadcast point placement approach in the following chapter.

LAMD implements a distributed relay-vehicle selection method, which chooses among the vehicles that have received an AM the most appropriate to ensure the full distribution of the alert message within the area of interest. This selection acts as a filter to avoid useless retransmissions. More precisely, on the advent of an alert message broadcasted by an RSU, a receiving vehicle individually decides whether it is eligible for rebroadcast by checking its location regarding the rebroadcast points. The closer a vehicle is from a rebroadcast point, the better candidate it is for a rebroadcast. This distributed distance-based priority scheme is implemented by enforcing different waiting times before rebroadcast to vehicles, i.e., linearly increase as a function of the distance vehicles have to their closest rebroadcast point. More precisely, vehicles sitting within a rebroadcast zone (which we name Candidate Relay (CR) vehicle) are invited to compute a waiting time as a function of their position with respect to the corresponding rebroadcast point within a time interval of $[0, W_{min} * T_s]$, where T_s is the time slot duration as defined in IEEE 802.11p and W_{min} is the minimum number of time slots.

To guard against road traffic situations where no vehicle is ideally positioned within a rebroadcast zone when an AM is broadcasted, as backup, vehicles outside the broadcast zones but at a distance less than $Rmax$ from the corresponding point (which we name Candidate Relay Backup (CRB) vehicle) are invited to wait for a time period within the $[W_{min} * T_s, W_{max} * T_s]$ (where W_{max} is the maximum number of slots) time interval before a rebroadcast attempt. When the vehicle with the shortest waiting time rebroadcasts the AM, all the CRs and CRBs vehicles that receive the AM cancel their rebroadcast attempt. For those that have already finished their waiting process and requested transmission from the MAC layer, this request is resumed using the MLMEX-CANCELTX" service primitive [Kuffermann 2014]. This avoids redundant retransmissions once a relay vehicle is notified that the AM was rebroadcasted by a better-positioned relay vehicle.

A last feature of LAMD is using a TTL (Time To Live) mechanism. Each AM conveys a TTL value initially set by the RSU when broadcasting the AM and then decremented at each rebroadcast. It hinders any rebroadcast when an AM message is received with a TTL equal to zero. The controller sets the initial value of the TTL for each RSU and allows extending the delivery coverage of an AM by allowing more than two-hops wireless paths. Indeed, in the case of a vehicle receiving an AM from a relay vehicle having no wireless connectivity with the network infrastructure (i.e., RSUs), the vehicle is elected as an opportunistic relay (that we call Candidate Relay Exceptional (CRE). If the received TTL is not null, it enters the waiting time process with a time period set as a function of its distance to the sending vehicle and rebroadcasts the AM. This TTL mechanism can be tuned on an RSU or an AM message basis, allowing the network controller to dynamically adjust the dissemination according to the wireless channel's sudden and transient degradation.

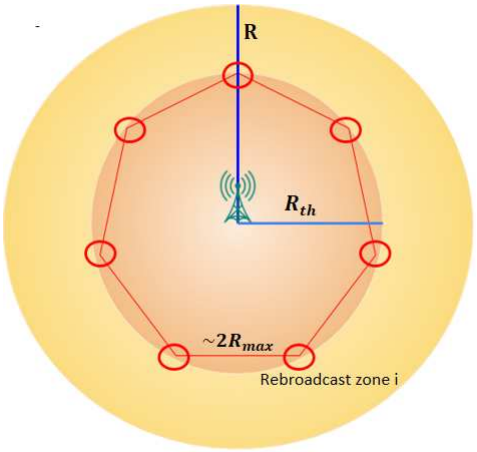


Figure 3.2: Rebroadcast zones

It can also be useful in case of ineffective or incomplete placement of rebroadcast points.

Figure 3.1 presents some examples of CR, CRB, and CRE vehicles with an illustration of their role in disseminating AMs.

3.3.2.1 Rebroadcast point placement

The number and the location of rebroadcast points and zones are inputs to the LAMD dissemination procedure. Therefore, different techniques could be considered for deriving, either a static placement or even a dynamic one, that adapts the placement to the wireless channel conditions (number of vehicles, weather, etc.). We describe hereafter a general and simple static placement method, which takes as input performance measures of each RSU. These can be derived from some site surveying or simulation studies.

Real experiments in [Xu 2017] show that the packet loss rate and delay significantly increase when the distance between vehicles and RSU exceeds a threshold R_{th}. We assume the controller can define a threshold distance R_{th} for each RSU. From this threshold, the controller builds a regular polygon with r equal sides ($5 \leq r \leq 17$), each with a length greater than $2R_{max} \pm 100m$, where R_{max} is the maximum transmission range of vehicles (as shown in figure 3.2). This ensures reduced interference between relay vehicles associated with two adjacent rebroadcast zones. For each polygon vertices, the controller derives the closest point on a roadside that falls within a distance of d_{max}. If such a point exists, it is added to the set of rebroadcast points.

3.3.3 Dissemination procedure

The proposed AM dissemination scheme combines relay selection and contention management. Algorithm 1 describes the behavior of a vehicle when receiving an AM. In this scheme, each node V_j maintains a list of rebroadcast zones in its

Algorithm 1: Alert message dissemination

Input:
AM : Alert Message
$P_i, i \in [1,k], d_{max}$: rebroadcast zones
V_j : current node, $Pos(V_j)$: its position, $Pos(V_j)$: its direction
S : Sender

```
AM.Reception()                                          /* AM received */
if S = RSU then
    P_i ← F(Pos(V_j), dir(V_j))    /* find the nearest rebroadcast zone */
    if d(P_i, V_j) ≤ d_max then
        V_j ← CR                          /* V_j is a Candidate Relay */
        T ← T_1                          /* Calculate time to wait T */
    else if d_max < d(P_i, V_j) ≤ R_max then
        V_j ← CRB                 /* V_j is a Candidate Relay Backup */
        T ← T_2
else
    if (AlreadyReceive (AM) = True) and (V_j = CR or V_j = CRB) then
        if d(P_i, S) < d(P_i, V_j) then
            Stop (t) and Discard()
    else if ( AlreadyReceive (AM) = True ) and (V_j = CRE) then
        Stop (t) and Discard()
    else if V_j ∉ {RSU} and TTL ≠ 0 then
        V_j ← CRE          /* V_j is a Candidate Relay Exceptional */
        T ← T_3
    else
        Discard ()
LaunchTimer (T)                                         /* start timer */
```

direction, shared by the controller during the handover. An AM is broadcasted with the sender's position, and a time to live TTL is used to control the rebroadcasts of vehicles located in gray zones, as explained above.

Let $P_i = F(pos(V_j), dir(V_j))$ a function which takes the GPS position of vehicle $pos(V_j) = (x_j, y_j)$ and its direction as parameters, and returns the closest rebroadcast zone in its direction.

As mentioned in Algorithm 1, if a vehicle V_j receives an AM, first, it checks the sender:

Step 1 [Lines 2-3] If the sender is an RSU, the node finds the closest rebroadcast point P_i from its location, and it checks its position according to the rebroadcast zone; otherwise, the message is received from a relay vehicle, go to step 4

Step 2 [Lines 4-6] If V_j is in the rebroadcast zone, the node concludes that it is a Candidate Relay (CR) and

then computes its waiting time T_1 as follows :

$$T_1 = W_{min} \frac{d(V_j, P_i)}{d_{max}} T_s \tag{3.1}$$

with $d(V_j, P_i)$ being the distance between vehicle V_j and rebroadcast point P_i, d_{max} is the radius of the rebroadcast zone, and W_{min} and T_s are respectively the minimum number of time slots and the time slot duration. As cited above, $T1$ lies between 0 and $W_{min} * T_s$. Go to step 7, otherwise, go to step 3

Step 3 [Lines 7-9] If V_j is beyond d_{max} but still within a distance R_{max} from the rebroadcast point, then, the node concludes that it is as Candidate Relay Backup (CRB) and computes the waiting time T_2 as follows :

$$T_2 = [W_{min} + (W_{max} - W_{min}) \frac{d(V_j, P_i)}{R_{max}}] T_s \tag{3.2}$$

where W_{max} is the maximum number of Time slots. T_2 is set between $W_{min} * T_s$ and $W_{max} * T_s$ to avoid collisions with nearby CR vehicles. This is particularly useful when the road traffic is low or sparse since, in case of no vehicle in a broadcast zone, a close-by vehicle is given the opportunity to rebroadcast the AM. Go to step 8, Otherwise discard the message

Step 4 [Lines 10-11] If V_j receives the same message another time, and V_j is a CR or a CRB go to step 5, otherwise, go to step 6

Step 5 [Lines 12-13] If V_j and the sender node are in the same rebroadcast zone, i.e., the distance between the rebroadcast point and the sender is less than the distance between the rebroadcast point and V_j, if the waiting process has not elapsed yet, it is stopped, and the AM is discarded. Otherwise, it cancels the MAC layer transmission request. Then, the procedure exits.

This condition ensures that once a CR has rebroadcasted an AM, no other CR from the same rebroadcast zone is allowed to retransmit the AM. It also avoids cases where a CR (or CRB) associated with a rebroadcast point is prevented from rebroadcasting if it receives a rebroadcast AM related to another rebroadcast point.

Step 6 [Lines 14-15] If V_j receives the same message another time, and V_j is a Candidate Relay Exceptional (CRE), stop the timer, the procedure exits, otherwise, go to Step 7

Step 7 [Lines 16-18] If V_j is not attached to any RSU and the $TTL > 0$, V_j is declared as CRE. In this case, the farthest vehicle from the sender, but still in its transmission range, has the highest priority to rebroadcast, so the waiting time T_3 is given by :

$$T_3 = W_{min}(1 - \frac{d(V_j, S)}{R_{max}})T_s \tag{3.3}$$

Otherwise, discard message

Step 8 [Line 21] Start the timer T

For completeness, it is worth noting that when V_j experiences a timeout, it updates the AM's TTL and then requests its transmission to the MAC layer.

3.3.4 Discussion

As explained above, LAMD focuses on message dissemination and takes a set of positioned rebroadcast points as input. Therefore, rebroadcast point placement is crucial to achieving the full performance of LAMD. But computing the optimal placement is difficult, as it requires an accurate propagation channel model on the area of interest, with varying characteristics and performance. One naive approach would resort to abusing the number of rebroadcast points, which raises scalability issues as the list of rebroadcast points that are reachable from an RSU must be provided to each vehicle during handover. It also adds contention between relay vehicles associated with nearby rebroadcast points as it is likely that they would be in the transmission range of each other.

Besides the TTL mechanism, which can be helpful for a non-optimal placement, we argue that thanks to the centralized control and the overall view of the network with additional information related to vehicles (their attributes, trip, etc.) brought from the cloud, effective dynamic placement can be devised and implemented by the controller. For instance, the controller could adjust the number and placement of rebroadcast points according to the number of available vehicles and their expected road trip (if available). Furthermore, at some rebroadcast points, the controller may instruct relay vehicles to transmit with a given transmission power or to direct the transmission towards a specific direction to increase the area of reach reliably. Such

instructions could even be applied to a specific AM, as they can be piggybacked with the AM. Finally, the controller could also learn from its experience, reported by past vehicles, by employing reinforcement learning techniques to adjust its placement according to the current context.

Another salient feature of LAMD is nominally allowing one retransmission by a CR or CRB per rebroadcast point. This implicitly assumes that this single transmission or the next AMs of the same road event are expected to be delivered successfully for any vehicle in the targeted gray area within a brief delay. Nominally, this is the case, but it may take some time to notify the event when vehicle density is low in some cases. When the controller identifies such a situation, this latter may proactively adjust AMs dissemination to the situation by, for instance, allowing more than one single retransmission per rebroadcast point. In other words, a CR or CBR would resume its AM rebroadcast attempt only if it receives a pre-defined number of rebroadcasts.

This work focuses on the downlink dissemination procedure (i.e., from the network infrastructure to the vehicles) since most performance gains lie there. In some scenarios, the vehicle may issue the AM that detects a road event. Then, the AM needs to be conveyed uplink toward the network infrastructure to benefit from the gains brought by the downlink dissemination. The uplink transmission in LAMD works as follows. If the vehicle has recently exchanged with an RSU (CAM or DENM messages or any other message on the SCH channels), it first invokes a unicast transmission to the RSU. If it fails or the vehicle has lost connectivity with its RSU, it broadcasts the AM stating that it is meant for uplink. Following the same logic as in downlink dissemination, any receiving vehicle computes its waiting time as a function of its distance to its RSU and then starts its waiting process as part of a rebroadcast attempt. The closer a vehicle is to its RSU, the shorter its waiting time is.

3.4 Performance evaluation

The goal is to assess the performance of LAMD regarding its ability: (1) to reach all the vehicles in the area of interest in a short time, (2) to avoid collisions and redundancy, and (3) to use network resources effectively. Also, our proposal is compared to the flooding technique [Ciccarese 2009], V2I broadcasts (with no V2V rebroadcasts) and the AddP [Oliveira 2017] protocol.

This section details the simulation environment, the performance metrics, and the obtained results.

3.4.1 Simulation setup

In the experiments, an event-based network simulator Netsim is used. Simulation of Urban Mobility (SUMO) bi-directionally coupled with Netsim as described in [Weber 2021] is used to make the most realistic simulations. SUMO is a microscopic road traffic simulator that allows creating a scenario by converting an existing map

Table 3.2: Simulation configuration parameters

Parameter	Value
Simulation time	500 s
AM generation start	10 s
AM generation rate	10 packets/s
AM packet size	1024 bytes
Propagation model	Nakagami m = 3
Slot time	10 μs
W_{min}	30
W_{max}	1023
d_{max}	16 m

or using one of the external tools provided by the SUMO project itself (for example, NETGEN or NETCONVERT) [Busanelli 2013]. Netsim includes implementations of IEEE 1609.4 and IEEE 802.11p communication standards. It also includes Basic Safety Message (BSM) handling and beaconing for cooperative awareness messages (CAMs).

In this work, we consider two urban scenarios with sparse, normal, and dense traffic conditions: in the first one, the roads form a symmetric Manhattan grid, while the second one corresponds to a part of a European-like city (namely, Toulouse, France) where the road structure is irregular. The vehicle density varied between 30 to 500 vehicles in the two scenarios. The duration of the simulation runs is 500 seconds. An average of 50 simulation runs is reported. The speed of vehicles is varied from 0 to 20 m/s, i.e., vehicles have different accelerations at different timestamps. We set the maximum transmission range of each vehicle to $R_{max} = 250m$.

In urban scenarios, signal shadowing effects heavily affect radio transmissions. The Netsim framework includes a log normal shadowing model. This model accurately captures large buildings' effect by blocking transmissions. For example, small walls block weak transmissions, while buildings only slightly hinder strong transmissions in the line of sight [Oliveira 2017]. We use this model in our experiments. In all simulated scenarios, there were two RSUs separated by a distance of 1 to 2 km. Each one handles the transmission of alert messages to vehicles in its transmission range. The other simulation parameters are shown in table 3.2.

3.4.1.1 Manhattan grid model

We generate a 12*4 Manhattan Grid scenario with a SUMO simulator covering an area of $2km \times 1.5km$ with road segments of 300 meters long and two lanes in each direction. This scenario is commonly used to evaluate network protocols in urban environments [de Sousa 2018, de Sousa 2015, Milojevic 2014, Araujo 2014, Garip 2015]. This type of network is called Manhattan Grid because its road

topology reproduces the Manhattan borough in New York City, USA. When reaching an intersection, a node can turn left, right, or go straight on, as in real life. So, in the Manhattan model, when the nodes reach the intersections, they move with 50% probability on the same street, 25% turning to left and 25% turning to the right, according to the scheme presented in [Meneguette 2014].

Inter-Vehicle Communication (IVC) algorithms easily handle data dissemination in Manhattan Grid scenarios since predesigned algorithms can efficiently recognize the orthogonal road topology. Therefore, Manhattan Grid scenarios are usually used in the first experiment to verify the validity of our algorithm. If an IVC algorithm cannot handle Manhattan Grid scenarios, it will be impossible to handle more complex ones. Moreover, due to the simple road topology of Manhattan Grid scenarios, some design issues can be clearly exposed [Tian 2016].

3.4.1.2 Toulouse city's scenario

In this scenario, we assume a cross-section of Toulouse, France, with an area equal to $2km \times 2km$. The map has been retrieved from Open Street Map (OSM) project. Here, the roads are characterized by an irregular shape, and there are junctions—note that the number of junctions is slightly higher than in the regular scenario.

3.4.2 Rebroadcast point placement

We proceed as follows. We first identify the gray zones for both considered city maps. Then, we apply the default placement method of section 3.3.2.1.
A wireless transmission's success depends on multiple factors: transmitter power, path loss, fading, receiver sensitivity, etc. Moreover, depending on the environment (highway, Urban, Obstructions, Line of sight), the transmission coverage of an RSU may vary widely, for example, from 100 to 700 meters, for the same transmission power as shown in [Gozálvez 2012, Böhm 2010]. On the other hand, road safety applications, especially those concerning cooperative awareness, require high reliability, i.e., a PDR between 90 and 95% [Boban 2017]. Therefore, for this study, the transmission coverage of a RSU is calculated as the maximum distance from the RSU where the PDR remains greater than 90% [Meireles 2010]. All the geographic areas that fall outside the transmission coverage of all RSUs are considered gray zones.
We have run several simulations to compute these zones where each RSU broadcasts a control message every 100 ms for 500s. After each packet, we record the distance of each vehicle from the sending RSU, the vehicles that received the message, and the packet status (success or error).

Figure 3.3 shows the computed Packet Delivery Ratio (PDR) as a function of the distance to the RSU for both scenarios. It can be observed that the PDR falls below 90% when the distance between the vehicle and the RSU is strictly greater than $R_{th}^{T} \simeq 600m$ for the Toulouse's scenario and $R_{th}^{M} \simeq 500m$ for the

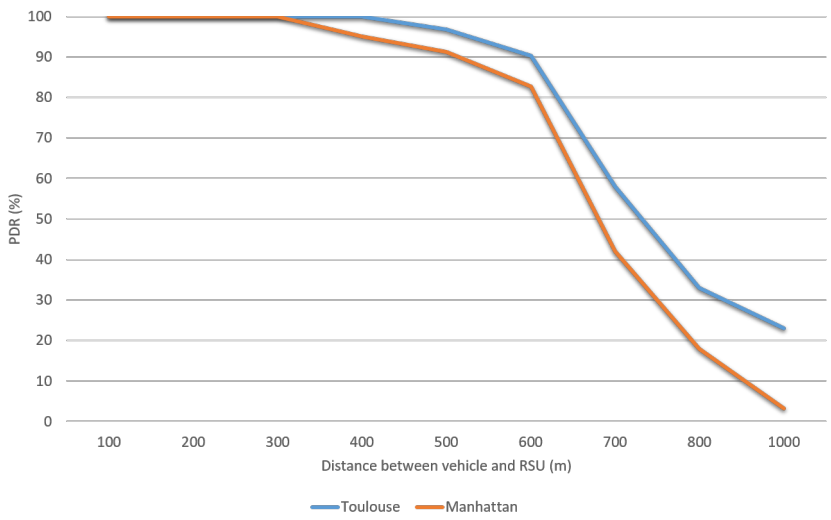


Figure 3.3: Packet Delivery Ratio (PDR) versus the distance between vehicles and RSU

Manhattan's scenario. From there, first, for the Toulouse's scenario, we build a regular octagon for each RSU with 8 sides of $459m(0.765R_{th}^{T}$ [Rezaei 2015]) in length and 8 vertices representing our rebroadcast points. Then, for Manhattan's scenario, we build a regular heptagon for each RSU with 7 sides of $434m$ lengths and 7 vertices representing the rebroadcast points in the Manhattan map. In fact, as mentioned above, we have placed the rebroadcast points at the edge of the road closest to each vertex. Also, we do not consider all the rebroadcast points as the RSUs are in the corners. As a result, we get 13 rebroadcasts zones in both Toulouse's scenario and Manhattan's scenario, as shown in Figure 3.4 and 3.5.

3.4.3 Performance metrics

The following three performance metrics are used to assess the effectiveness of LAMD in disseminating AMs.

- **Information Coverage (IC):** computed as the total number of vehicles that successfully receive (NumVehRecMsg) an AM at the end of the simulation divided by the number of vehicles (TotalNumVeh) averaged on all generated AMs. This metric shows how successful the dissemination is after a decent period of time.

$$IC = \frac{NumVehRecMsg}{TotalNumVeh} \tag{3.4}$$

- **Packet Delivery Ratio (PDR):** is the ratio between the number of AMs (NumRecMsg) successfully received by each vehicle and all transmitted AMs (NumTransMsg) during the vehicle lifetime within the area of interest averaged on all the vehicles of the area of interest. This metric is important

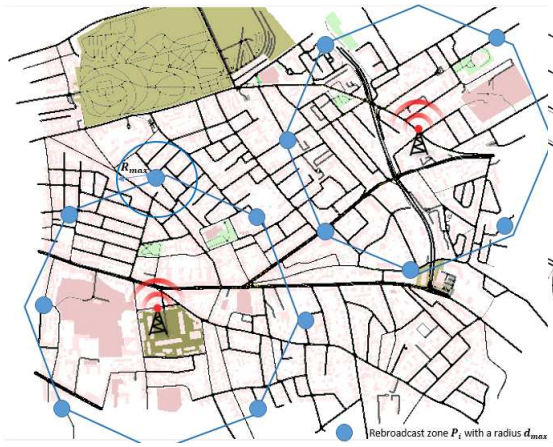


Figure 3.4: Toulouse's city scenario

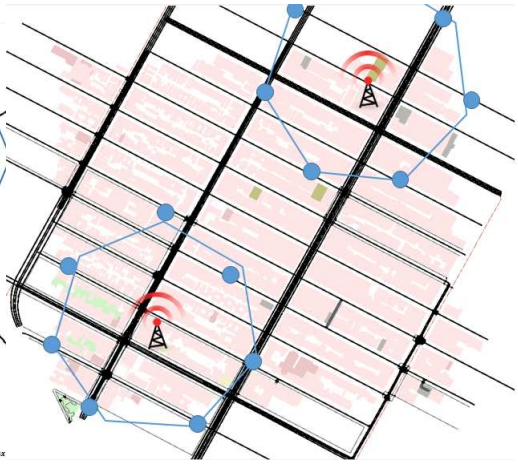


Figure 3.5: Manhattan's city scenario

since it measures the sustained effectiveness of the dissemination.

$$PDR = \frac{NumRecMsg}{NumTransMsg} \tag{3.5}$$

- **Dissemination delay:** The dissemination delay is the total time required to deliver the AM to all the vehicles in the area of interest that receive the AMs. The vehicles that do not receive any AM are excluded from the computation. This metric measures how fast the dissemination can reach the vehicles within the area of interest.

The following two performance metrics measure the network overhead induced by the dissemination in terms of collisions and redundant successful transmissions.

- **Collision Ratio (CR):** The collision ratio is the percentage of MAC collisions (NumCollisionPkts) divided by the number of packets sent (NumSentPkts) computed over the simulation duration.

$$CR = \frac{NumCollisionPkts}{NumSentPkts} \tag{3.6}$$

- **Redundancy Rate (RR):** The average number of AM rebroadcasts or duplicate transmissions (i.e., NumDuplicPkts) out of all sourced (generated) AM (NumSPkts) [Panichpapiboon 2011].

$$RR = \frac{NumDuplicPkts}{NumSPkts} \tag{3.7}$$

3.4.4 Performance analysis

This section details the performance results for Toulouse and Manhattan scenarios under various vehicle densities.

3.4.4.1 Effectiveness of LAMD dissemination procedure

i. Information Coverage

Figures 3.6 presents the Information Coverage as a function of vehicle density for LAMD, flooding, V2I, and AddP when applied to the scenario of Toulouse city. The results show that starting from a vehicle density of 225 vehicles, LAMD achieves full vehicle coverage at the end of simulation time. In comparison, AddP achieves full coverage with the same vehicle mobility, starting from a vehicle density of 420 vehicles (almost twice the density with LAMD). This is because LAMD combines V2I and V2V transmissions while AddP relies exclusively on V2V transmissions. These latter depend on the vehicle density as vehicles must be in mutual radio coverage to support a V2V transmission.
For the considered experiments, the flooding dissemination never achieves a full-coverage even under high densities, despite all vehicles being allowed to rebroadcast an AM. Worse, when the density increases, the IC decreases. In fact, the increase in vehicle density increases transmission collisions, and redundant rebroadcasts. Both consume the scarce radio resources in place of constructive rebroadcasts. This is a typical illustration of the negative effect broadcast storm problem cited above.
The IC of V2I also decreases when vehicle density increases. By increasing density, more vehicles are likely to be located outside of the RSU coverage. The results show that with the vehicle mobility generated by SUMO on average, more than 20% of the vehicles do not fall within reach of an RSU for the entire simulation period (500s) or their lifetime. Despite this important portion of far away from vehicles, even under very low vehicle densities where the presence of vehicles within or around rebroadcast zones is not guaranteed, LAMD achieves an IC of more than 92%. Moreover, as shown in Figure 3.7, which plots the IC as a function of time for low vehicle density scenarios ($35, 75, 225$ vehicles), the maximum achievable IC is reached in less than $500ms$.

As shown in figure 3.8, the evolution of IC as a function of vehicle density for Manhattan's scenario is quite similar to Toulouse's scenario. However, we notice an expected slight difference because the road layout is different. Notably, with LAMD, full vehicle coverage is achieved starting from a lower density of vehicles, i.e., with a density of 125 vehicles (vs. 225 vehicles for Toulouse's scenario).

ii. Packet Delivery Ratio

Figure 3.9 and Figure 3.10 present the PDR for all dissemination procedures, respectively, for Toulouse's city scenario and Manhattan's city scenario under different vehicle densities. The results show that LAMD sustainably delivers AM messages to almost all vehicles (with a PDR of 96%). Notably, in comparison, AddP exhibits a significantly lower PDR (around 70%). This shows that LAMD distributes AMs more effectively by presumably selecting more suitably relay vehicles and inducing fewer packet losses over the overall simulation period. This comparison also holds with flooding, whose poor packet delivery ratio is caused by the collisions generated by the broadcast storm.

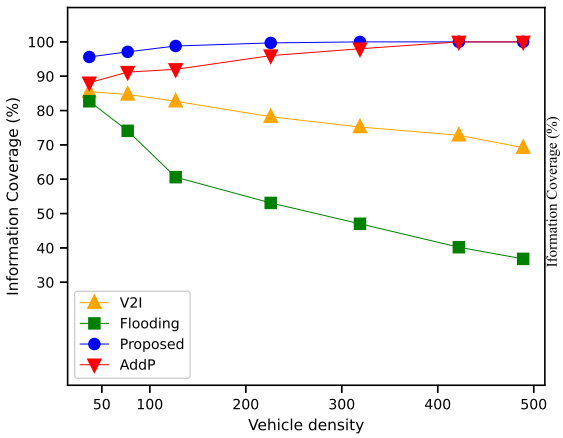


Figure 3.6: Information coverage - Toulouse

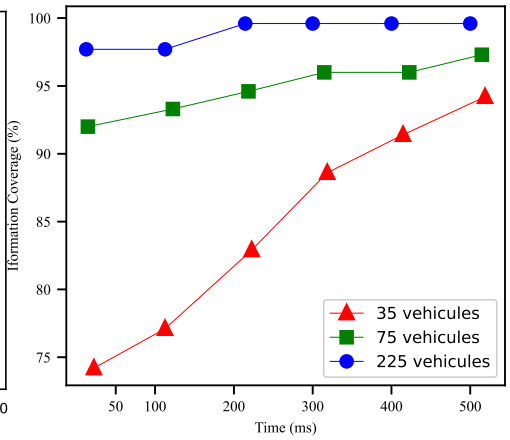


Figure 3.7: Information coverage - Toulouse

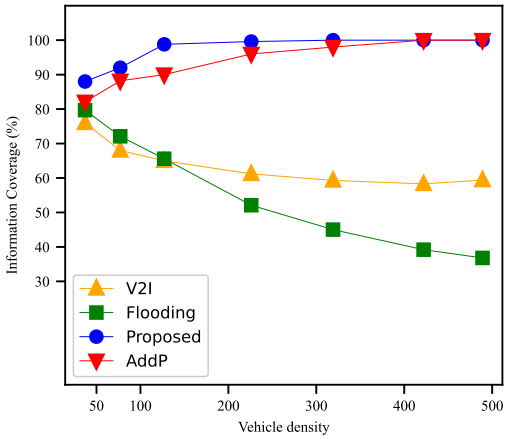


Figure 3.8: Information coverage - Manhattan

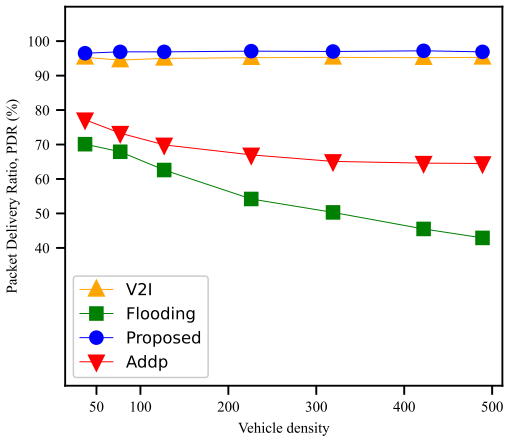


Figure 3.9: Packet Delivery Ratio - Toulouse

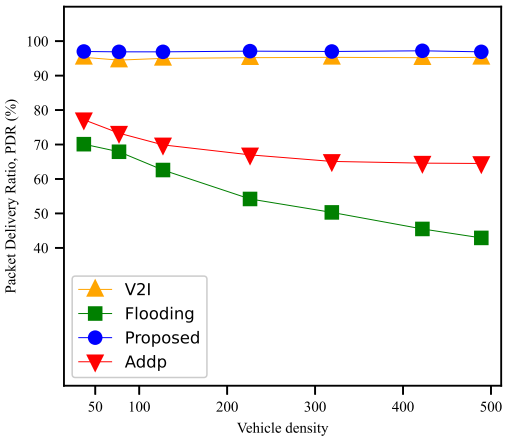


Figure 3.10: Packet Delivery Ratio - Manhattan

iii. Dissemination Delay

Figure 3.11 and Figure 3.12 show the dissemination delays versus vehicle density respectively in Toulouse's city and Manhattan's city scenario. The results show that the dissemination delays with LAMD and AddP are significantly lower than the delays observed with Flooding and V2I approaches, especially for high traffic densities.

The results show that LAMD delivers the first copy of an AM to all vehicles within a few AM periods. For high vehicle densities, relay vehicles always exist, which quickly rebroadcast the AM to the vehicles outside the RSUs coverage, leading to a dissemination delay below twice the AM period. For low vehicle densities, fewer vehicles are available in the network, the presence of vehicles around each rebroadcast zone is hence not guaranteed for the first AMs. Nevertheless, LAMD can still deliver an AM to all the vehicles within 5 AM periods (i.e., $500ms$). For instance, for a vehicle density of 30 vehicles, the total time required is $315.28ms$ in Manhattan's city scenario, with over 75% of vehicles served within $100ms$.

AddP exhibits similar results thanks to its relay selection method, which favors far away relay vehicles with the highest neighborhood. This helps speeding the full dissemination of AMs.

Even if all vehicles are allowed to rebroadcast an AM with the potential of fastly reaching all the vehicles, the flooding dissemination procedure suffers from the effect of the broadcast storm. AM rebroadcasts experience collisions and increased contention, delaying their successful delivery.

For the V2I approach, it takes quite a long time to reach all the vehicles in the area of interest, as these latter need to be under the coverage of an RSU to receive an AM. As the average speed of vehicles is between 7 and 13 m/s, i.e., the vehicle can only move approximately 1m in $100ms$. So, it may take some time before a vehicle comes under the coverage of an RSU. This is the reason why V2I offers by

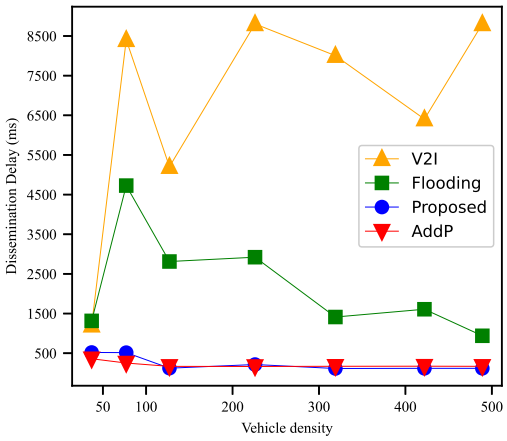


Figure 3.11: Dissemination delay - Toulouse

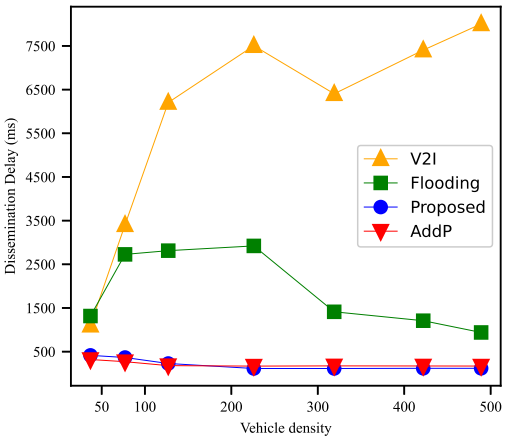


Figure 3.12: Dissemination delay - Manhattan

far the worst dissemination delays.

3.4.4.2 LAMD network overhead

i. Collision Ratio

Figure 3.13 and Figure 3.14 show the collision ratio as a function of vehicle density respectively for Toulouse's city scenario and Manhattan's city scenario. The results are similar in both scenarios. It is worth noting that even for high vehicle densities, the collision ratio remains very low with LAMD (about 0.001%). In comparison, we observe that the collision ratio reaches 59% for flooding and ranges from 22% to 34% for AddP (yet AddP also relies on some form of relay selection to limit contention to the wireless medium).

This is obviously an important advantage of LAMD since radio resources are saved and made available to other AMs or CAM messages. This is achieved thanks to the selection method, which limits the set of candidate relays to those close to rebroadcast points and then further reduces rebroadcast contention by assigning different waiting periods to candidate relays. However, some rare collisions may still occur, either because of a simultaneous transmission from two or more nearby candidate relays associated to different rebroadcast points, or the back-off procedure of the CSMA/CA technique, which if the wireless medium was sensed busy, introduces a random back-off time when the vehicle accesses the medium.

ii. Redundancy Ratio

Figure 3.15 and Figure 3.16 show the number of duplicated AM transmissions for all techniques successively in Toulouse's city scenario and Manhattan's city scenario under different vehicle densities. As expected, Flooding is the protocol with the highest redundancy because no mechanism limits AM rebroadcasts. Compared to other techniques, LAMD induces the smallest redundancy ratio (below 50%).

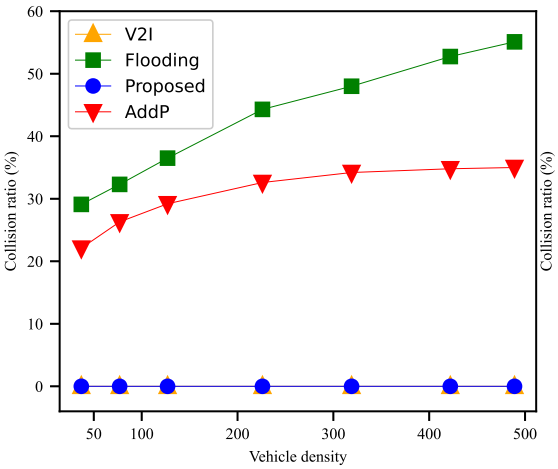


Figure 3.13: Collision ratio - Toulouse

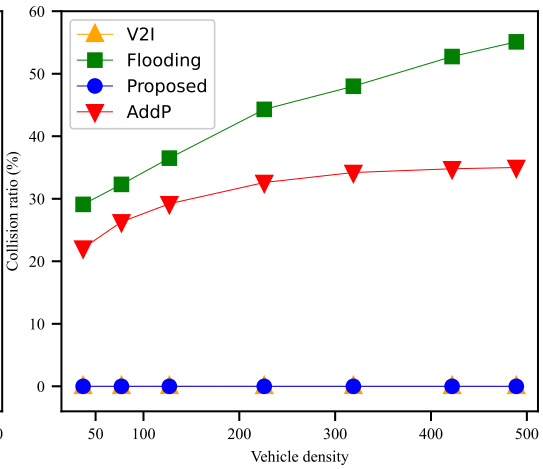


Figure 3.14: Collision ratio - Manhattan

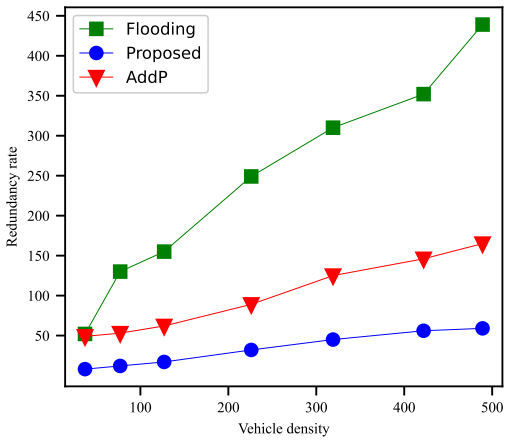


Figure 3.15: Redundancy ratio - Toulouse

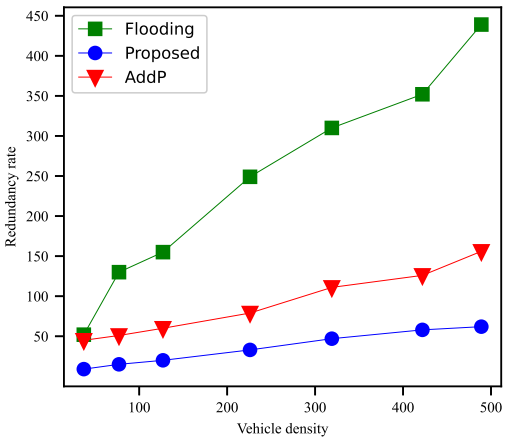


Figure 3.16: Redundancy ratio - Manhattan

The redundancy gain with respect to the other dissemination schemes increases as the vehicle density increases and reaches one third less redundancy than AddP (as explained above, also uses a relay selection method that reduces transmission redundancy).
This low redundancy is achieved thanks to the ability of LAMD to correctly select only a small set of relay vehicles to fully disseminate an AM, and also to the fact that a candidate relay that observes an AM rebroadcast resumes its rebroadcast attempt.

3.5 Conclusion

This chapter proposes an alert message dissemination procedure based on vehicle locations to provide high coverage, fast delivery, and minimal network overhead. The main peculiarities of our solution are: First, the combination of V2I transmissions and V2V re-transmissions, where first, RSUs broadcast alert messages, and then, some selected relay vehicles rebroadcast the messages after a personalized waiting time. Second, the definition of broadcast zones by the SDN controller where relay vehicles may rebroadcast the message to reach all vehicles that are located in the area of interest, particularly those in gray areas. To that end, we have proposed a static default method that computes rebroadcast points. The performance results prove the effectiveness of our procedure by providing a high delivery ratio, limiting packet collisions, minimizing redundancy, and ensuring rapid dissemination and high information coverage.

In this chapter, we have detailed and highlighted the advantages of our proposed dissemination procedure. In the next chapter, we will focus on a new reinforcement learning technique of rebroadcast point placement that adapts to radio channel conditions, the actual locations of the gray zones, and road traffic conditions.

Chapter 4

Q-learning-based Rebroadcast Zones Placement For An Effective Alert Message Dissemination

4.1 Introduction

The Alert message dissemination procedure is the basis of the approach proposed in this thesis. First, we have shown this approach's advantages and benefits. Then, we have highlighted issues that need improvement to make this scheme more efficient and accurate.

Among these issues, we mention the placement of the rebroadcast zones/points. Indeed, as described in the previous chapter, the network controller computes rebroadcast zones (or rebroadcast points), and only the vehicles that sit in their close vicinity (i.e., rebroadcast zones) are allowed to rebroadcast an AM. The computation of these rebroadcast zones takes advantage of the centralized view that the network controller builds on the wireless network coverage, radio propagation environment, as well as of the information related to road traffic and vehicles (their characteristics and potentially their road trip).

In this chapter, we propose to use a new dynamic approach, which consists of adjusting the placement of the rebroadcast zones/points according to the evolutions of the gray zones due to the regular changes in the traffic conditions. We first describe the problem and the method adopted. Then, we present our placement algorithm based on reinforcement learning. Finally, we consider the same map of the city of Toulouse to evaluate the performance gains of our method compared to the first default method proposed in the previous chapter.

The rest of the chapter is organized as follows: Section 4.2 presents an overview of existing works in the literature. Section 4.3 describes the problem. Section 4.4 presents an overview of reinforcement learning. Section 4.5 details the proposed method. Section 4.6 evaluates and discusses our simulation results. Section 4.7 concludes the chapter.

4.2 Related work

In the literature, no work directly deals with rebroadcast zone placement problems because the relay selection technique initially proposed in the previous chapter differs significantly from other previous works, considering only the location of vehicles (so its presence or not in one of the rebroadcast zones). In the same way, service placement is an entirely different subject that should not be confused with the rebroadcast zone placement proposed in this work. Another problem close to our work is the optimal RSU placement for vehicular networks. Instead, RSU placement is a location optimization problem that aims to find the optimal positions to cover all nodes while considering several constraints, such as deployment cost and application requirements.

In [Aslam 2012, Cavalcante 2012, Sun 2010, Liu 2013], the authors consider the problem of optimal RSU placement in urban areas. Their goal is to place RSUs optimally to enhance connectivity. They consider intersections as the best locations in urban scenarios since the density in junctions is usually higher, and information can disseminate in all directions. In [Aslam 2012], Aslam and al. propose two optimization methods to resolve the problem of minimizing the average reporting delay of safety messages propagation for a given number of RSUs and area coverage: The Binary Integer Programming (BIP) method using a branch and bound approach to find the best delay to broadcast in the entire region and the Balloon Expansion Heuristic (BEH) method using balloon expansion analogy to find the best delay in each route. In [Cavalcante 2012], the goal is to improve data dissemination in an urban area. They use a genetic heuristic to solve the Maximum coverage with Time Threshold Problem (MCTTP [Trullols 2009]) that models the location problem and gets the best RSUs placement to cover the maximum number of vehicles on the road. In [Sun 2010], a cost-efficient RSUs deployment scheme is proposed, where authors aim to update security certificates within a brief delay using the deployed RSUs. In this scheme, the authors suppose that each vehicle can exchange with RSUs in bounded time, and the overhead time is used to adjust routes to update certificates. The problem is modeled as a set-covering problem and solved by the polynomial-time approximation algorithm called "Greed Set Cover" [Aslam 2011]. In [Liu 2013], Liu and al. model the road network as a weighted undirected graph where each edge represents the average passing time on the corresponding road. They use the depth-first traversal algorithm to get optimal placement to deploy RSUs. In [Mehar 2015], the authors use a genetic algorithm and Dijkstra algorithm to reduce the number of RSUs based on the deliverance time applications requirement and the deployment cost. However, all those works consider only V2I communications, which introduce a significant delay for vehicles not available to drive toward the RSUs.

Unlike the existing works, our work makes contributions by considering the dynamic rebroadcast points placement problem with no need to deploy new RSUs

or any network components (so, no additional cost). Furthermore, developing an RL-based solution provides an optimal number and placement of rebroadcast zones with a low delay while keeping the server resource utilization low.

4.3 Problem description

We consider the location-based alert message dissemination (LAMD) scheme detailed in the previous chapter; starting from the AM broadcast from the network infrastructure, each RSU selects a set of relay vehicles that can re-transmit the AM to reach vehicles in areas that are poorly or not covered (gray zones) by the transmissions from the network infrastructure. Indeed, a vehicle that receives an AM individually decides whether it is eligible for rebroadcast by checking its location with respect to pre-defined rebroadcast points sent to the vehicle during handover. The closer a vehicle is from a rebroadcast point, the higher its priority is to rebroadcast the AM. Therefore, rebroadcast zones' placement is crucial for the efficiency of such dissemination schemes, as the selection of rebroadcasters is based on those zones.

We consider a rebroadcast zones placement problem in a geographic map with gray areas having poor or no coverage. Given a geographical map represented by a matrix of small squares of $40 \times 40m^2$, we can distinguish gray areas, road areas, and non-road areas in each square. While considering the gray square's position, the problem is to find the optimal number and placement of rebroadcast zones. The aim is to select the region's optimal rebroadcast squares that maximize the gray squares' total coverage. The number of gray zones can be dynamic by considering several factors (weather, traffic conditions, etc.). The default method with fixed rebroadcast points is insufficient for a vehicular network's mobile and dynamic scenario. Therefore, the real-time environment must be considered while placing rebroadcast points. With this goal, we improve the previous default solution by using a Reinforcement Learning (RL) rebroadcast points placement algorithm adapted to the changing vehicular environment. Our solution uses a classic model-free Q-learning algorithm that finds the optimal rebroadcast zones based on gained experience from interactions with the environment.

4.4 Background

4.4.1 Reinforcement Learning (RL)

RL is the science of decision-making or the optimal way of making decisions in a defined *environment* (e.g., a Freecell game) with which an *agent* (what we will program an RL algorithm into. e.g., a player) can interact. The agent seeks to select *actions* (e.g., to play moves) to maximize a *reward* throughout an episode (e.g., a game). Each time the agent chooses an action, the environment changes, and the agent receives a reward and moves to a new state where new actions are possible. The environment is modeled mathematically in RL by adopting Markov

Decision Process (MDP) formalism, and the agent as a decision rule (or policy) that selects actions based on the history of its previous interactions and knowledge about the current state of the environment.

A key challenge in RL is that the agent has limited or no prior knowledge about the impacts of its actions on the environment: it can only collect samples from the environment by taking actions and observing their effects. The RL algorithm performance is usually measured regarding how many samples are required from the environment to learn a good policy or the intensity of the "errors" it makes during learning. To achieve optimal performance, it must handle two objectives simultaneously; learn *how the environment behaves* and *how to act optimally* while interacting with the environment. The first objective is called exploration, and the second is exploitation. Balancing these objectives is referred to as the *exploration-exploitation dilemma*. RL algorithms must use a similar dilemma to maximize rewards in an unknown environment with limited resources.

At every time step, the environment is described by a state variable s belonging to a state set S. The complexity of exploration is related to the size of the set S, which can be its cardinality (if it is finite) or some notion of dimension. For instance, if S is finite and no prior information is given about the environment, an agent must visit every reachable state $s \in S$ to learn a good policy. Otherwise, it might miss a state with very high rewards.

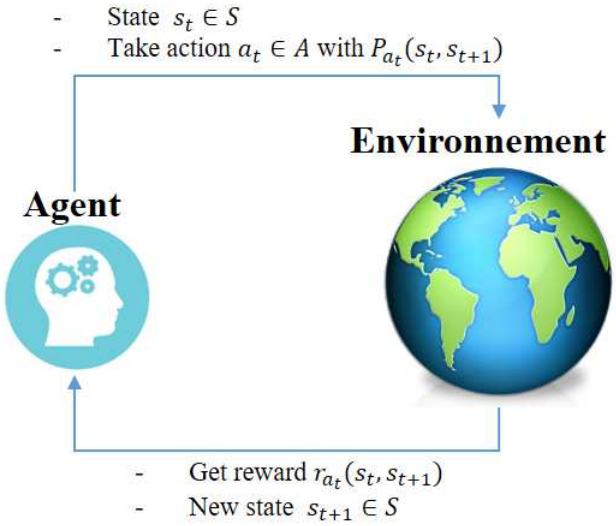


Figure 4.1: Abstract Markov Decision Process model

4.4.2 Markov Decision Process (MDP)

The environment with which an agent interacts is modeled as a Markov decision process (MDP) [Puterman 2014]. An MDP is defined as a (S, A, P, r) tuples, where S stands for the set of possible states, A_{s_t} is the set of possible actions from state $s_t \in S$ to $s_{t+1} \in S$, $P_a(s_t, s_{t+1})$ is the probability to go from state s_t to s_{t+1} performing an action $a \in A$, $r_a(s_t, s_{t+1})$ is the immediate reward, earned from the transition from state s_t to state s_{t+1} by performing an action a, as seen in Figure 4.1. The reward is a stimulus, positive or negative, that the agent seeks to maximize in the long term.

The decision policy π is a function that maps state to action (the agent's brain); basically, a policy function says what action to perform in each state. The ultimate objective with an MDP lies in finding the optimal policy π^* which specifies the correct action $\pi(s) \in A$ to perform in each state $s \in S$, which maximizes the sum of reward. In other words, we would like our agent to learn a function that enables it to map S to A ($\pi : S \leftarrow A$).

A *Q-value function* determines how good it is to end up in a particular state with a policy (π). This function gives a score according to the sum of predicted future rewards, also called the *expected return*. Accurately, the expected return can only be determined if the agent is deterministic in choosing an action and if the agent uses the same policy through an episode. Mathematically, a policy is just a mapping between states and actions on the one hand and the probability of selecting this action in this state on the other hand. A Q-value function is often denoted by $Q(s)$.

4.4.3 Bellman optimality equation

Bellman equation is the basic block of solving reinforcement learning and is omnipresent in RL. It helps us to solve MDP. To solve means finding the *optimal policy* that has the best *expected return* denoted $Q^*(s)$ in all states. This policy can be calculated using the Bellman optimality equation :

$$Q^*(s_t) = \max_{a \in A(s_t)} Q_{\pi^*}(s_t, a) = \max_{a \in A(s_t)} \sum_{s_{t+1}, r} p(s_{t+1}, r | s_t, a)[r + Q^*(s_{t+1})] \tag{4.1}$$

Where

- $Q^*(s_t)$ is the Q-value function of state s under optimal policy
- $Q_{\pi^*}(s_t, a)$ is the Q-value of taking action a in state s_t under optimal policy
- $p(s_{t+1}, r | s_t, a)$ is the probability of an action a in state s_t to end in state s_{t+1} with reward r
- $r + Q^*(s_{t+1})$ is the recursive return of s_t under optimal policy
- $A(s_t)$ is the set of all possible actions in state s_t

Using this equation, we can write down a system and solve it. The main problem is that the system has as many equations and unknowns as there are different states. For many practical problems, there are too many states, and the system is too complex to be solved. A solution is to use another approach to approximate the optimal policy, a policy that is good enough to play the game near optimally [Castiaux 2021].

4.4.4 Policy iteration

One of those different approaches to solving an RL problem is via experience. First, the agent is placed in the environment and simulates a full episode according to a policy π. When the episode is finished, the return Q-value is determined and can be back-propagated to all state-action pairs throughout this episode. The operation must be repeated in many episodes in order to gain knowledge and improve the approximation of Q_π. This stage is the evaluation of Q_π for π.

When the evaluation of Q_π^i gets precise enough, it is possible to determine a new $\pi^{'}$ policy that always selects the best action according to Q_π. This stage is known as the improvement of π. It is possible to keep improving a policy until it reaches a local maximum, a local best policy, by continuously evaluating then improving the current policy.

4.5 Proposed RL-based rebroadcast zones placement

This section details the features of the proposed RL algorithm to solve the rebroadcast zone placement problem. We first summarize the environment representation, the state space, reward function, and the action space used in our RL framework.

4.5.1 Environment, States, Actions, and Rewards

- *Environment*: Our environment is represented by the geographical map covered by the SDN controller as a matrix of size $L \times M$ (L and M depend on the size of the geographic map) of small squares of $40 \times 40m^2$ in order to manage the quality of reception links in each small square area "Figure 4.2". The area of the square could be enlarged or reduced according to each zone of interest covered by the controller (urban area with many intersections/buildings or highways). Five values are assigned to define each square's type as mentioned in table 4.1.

- *State Space S*: The state space contains the number and positions of rebroadcast zones, e.g., for state s at iteration t, $s_t = \{N, (x_1, y_1), ..., (x_N, y_N)\}$, where, N is the number of rebroadcast zones at the current state and, (x_i, y_i) is the coordinate of the i^{th} rebroadcast zone at the current state.

Figure 4.2: Environment representation

- *Action Space A*: The number of rebroadcast zones can be incremented by adding (A(x_j, y_j)) a new rebroadcast zone $j(x_j, y_j)$, decremented by removing an existing rebroadcast zone, or maintained ($\overline{A}$). The position of each rebroadcast zone in a current state can be deleted (DL), maintained (M), or moved up (U), down (D), right (R), or left (L). The number and positions of rebroadcast zones are updated in each iteration by performing one of the following actions.

$$s_{t+1} \xleftarrow{N \times a^i \in \{DL,M,U,D,R,L\} \cup a \in \{A(x_j,y_j),\overline{A}\}} s_t$$

  We add a constraint to avoid collisions between two adjacent rebroadcast zones. The distance between two rebroadcast zones must be greater than or equal to a threshold φ that depends on a vehicle's average coverage in the rebroadcast zone.

- *Reward Function*: The reward for each square that can be elected as a rebroadcast zone (value: "0") is calculated according to its communication range in meters (the communication range of a vehicle in the center of this zone), where we can estimate the number of gray squares covered by this area. So, for a square i($value = 0$), the reward is: r_i = number of gray squares covered by i. For state $s_t = \{\cup_{i=1}^{N}(x_i, y_i)\}$ at iteration t, the reward r_t is the sum of rewards of all elected rebroadcast zones: $r_t = \sum_{i=1}^{N} r_i$.
  A gray square already covered by an elected rebroadcast zone cannot be

Table 4.1: Squares' values

Square's type	Value
Road zone could be elected as rebroadcast zone	0
Non-road zone or zone X which could not be elected as rebroadcast zone	1
Gray zone	0.5
Gray zone covered by a rebroadcast zone elected	1.5
Rebroadcast zone elected	2

covered by another rebroadcast zone, that's why we change its value from 0.5 to 1.5 (in other words, a newly elected rebroadcast zone cannot consider a square of value 1.5 as a gray zone).

The objective is to select the optimal rebroadcast zones in the region that maximize the total coverage of the gray areas(squares = 0.5).

4.5.2 Q-learning and problem formulation

There are many practical scenarios, such as the placement problem studied in this work, for which the transition probability $P_a(s_t, s_{t+1})$ is unknown, which makes it difficult to evaluate the policy. Q-learning is an effective and popular algorithm for learning from delayed reinforcement to determine an optimal policy without the transition probability. It is model-free reinforcement learning that allows agents to learn how to act optimally in Markovian domains by experiencing the consequences of their actions without requiring maps of these domains.

Following the above system description, we can model the problem as a discrete-state MDP, where an agent (SDN Controller) in a state s_t takes action $a_t \in A$ and transitions to another state s_{t+1}. As a result of the execution of this action, the environment returns a rebroadcast zone's position dependant reward r_t, which allows the local update of a Q-value, $Q(s_t, a_t)$, indicating the appropriateness of selecting action a_t in-state s_t.The Q-value is computed according to the rule [Sutton 2018]:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_t + \gamma \max_{a \in A} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \tag{4.2}$$

where α quantifies to what extent the newly acquired information will override the old information. An agent with $\alpha = 0$ will learn nothing, while with $\alpha = 1$ it would consider only the most recent information, and $\gamma \in [0, 1]$ is the discount factor that determines the current value of the future state costs.

The RL algorithm proposes to converge to an optimum solution in terms of the rebroadcast zone's number and position. Authors in [Watkins 1992] proved that Q-learning converges to the optimum action-values with probability "1" as long as all actions are repeatedly sampled in all states and the action-value pairs are represented discreetly. First, the algorithm randomly selects N rebroadcast zones (N is approximately calculated as the ratio of the total number of gray squares

and the average number of gray squares that a rebroadcast zone can cover). Then, the greedy policy regarding the Q-values tries to exploit continuously. During the learning phase, the agent selects the corresponding action based on the ε-greedy policy, i.e., it selects with probability $1-\varepsilon$ the action associated with the maximum Q-value, and with probability, ε selects less frequently ($\varepsilon = 0.1$) a random action (here, $y \in [0, 1]$ is a random variable):

$$a_t = \begin{cases} \arg\max_{a \in A} Q(s_t, a_t), & \text{if } y > \varepsilon \\ rand(A), & \text{otherwise} \end{cases} \tag{4.3}$$

This means that the controller uses the optimum Q_value 90% of the time and makes exploratory actions 10% to gain new experience. This balancing between exploitation and exploration can guarantee convergence and often good performance. This is because the controller will have explored all possible actions and avoided local minima [Sutton 2018].

4.5.3 Q-learning algorithm

Algorithm 2: Q-learning-based Rebroadcast zone placement

Input:
$L \times M$ Environment Matrix
N: Initial number of rebroadcast zones
Output:
$L \times M$ Environment Matrix with optimal rebroadcast zones placement and gray zones covered

Initialize the first state with N positions of rebroadcast zones randomly selected, $Q_0(s, a) = 0, \forall s \in S$ and $\forall a \in A$ at iteration $t = 0$
while *Learning* **do**
  Visit state $s_t = (\{\cup_{i=1}^{N}(x_i, y_i)\})$
  for $(x_i, y_i) \in s_t$ **do**
    Select an action a^i using ε-greedy rule in (4.3)
    Update The values of Environment Matrix elements
    Calculate the reward r_t^i
    Observe next state s_{t+1}^i
    Update the Q-value $Q(s_t^i, a^i)$ from (4.2)
  Select an action a `/* to increment or not` N `*/`
  if $a = A(x_j, y_j)$ `/* add a new broadcast zone j */` **then**
    Update The values of Environment Matrix elements
    Calculate the reward r_t
    Observe next state s_{t+1}
    Update the Q-value $Q(s_t, a)$

As described above, the environment is represented as a $L \times M$ matrix of $40 \times$

$40m^2$ small squares. Indeed, we initialize matrix elements with only one of the three first values in Table 4.1 ("0", "1", and "0.5"). Then, at each iteration, the value of elements "0" and "0.5" can be updated respectively to rebroadcast zone elected ("2") and gray zone already covered by a rebroadcast zone elected ("1.5") and vice versa. A gray zone (value = "0.5") can only be covered by one rebroadcast zone (value = "2"). In order to limit the number of rebroadcast zones and the collision between two adjacent rebroadcast zones, we add a constraint when performing actions. The minimum distance between two adjacent rebroadcast zones is 5 squares ($200m$, knowing that the maximum coverage of a vehicle in the center of a rebroadcast area is $250m$ (≈ 6 squares)).

The Q-Learning algorithm is described in Algorithm 2. The set of gray squares is taken as input as we assume that the controller has a prior and updated vision of the quality of links in each road segment. This can be achieved thanks to some wireless site surveying, simulations, or predictions as in [Benrhaiem 2020]. The first step of the algorithm randomly selects N feasible rebroadcast points. After that, at each iteration t, the position of each rebroadcast zone ((x_i, y_i), for all $i \in N$) and then the number of rebroadcast zones make an exploratory move with probability ε or picks the best-known action to date (highest Q-value) with probability $1 - \varepsilon$. The algorithm explores different states during the learning phase (a fixed simulation/iteration run) to find the optimal policy that maximizes the expected action-value function (Q-value) and, hence, the total coverage of gray zones. The distance between a new candidate and other elected rebroadcast zones is always checked before choosing and executing an action.

If n is the number of states and e is the number of actions (6 in our case), empirical results in [Koenig 1992] show that the Q-learning algorithm ends in $O(en)$ steps; hence, the computational complexity of our algorithm is $O(6nN)$, where N is the number of rebroadcast points. Furthermore, the complexity is further reduced if the domain has additional properties, such as constraints between two adjacent rebroadcast points [Koenig 1992].

In conclusion, our method exhibits the following advantages while using decent memory and computational resources compared to traditional approaches. First, RL algorithms are applicable to environments where no prior information, assumptions, or requirements about the region considered are available (e.g., heuristic methods wait to receive triggers before taking a decision, and are effective for enabling agents to explore unrecognized state spaces and learn information pertinent to performing tasks. Second, learning-based methods are used when there is a possibility of further optimizing the system based on learning from the history of the triggers mentioned above). Third, the decision-making process of Q-placement is on-demand. Unlike traditional algorithms with fixed optimization levels, Q-learning lets one decide the optimization level. This is achieved by tuning the number of iterations the algorithm runs. It is a desirable feature because the controller can determine precisely how much computation power it commits to achieving a certain performance level. This flexibility is essential, especially in situations where the controller is time-constrained. Moreover, the Q-learning

algorithm is fully compatible with a centralized architecture, e.g., the Q-learning algorithm itself can be regarded as an application running on the controller. Finally, all the information needed by Q-learning is collected during routine network status updates between the controller and vehicles [Zhang 2018].

4.6 Performance analysis

This section details the performance results of our algorithm and compares its performance to the proposed default one using the location-based dissemination procedure described in the previous chapter. First, we describe the experimental setup before diving into the performance analysis of our placement method taken alone and combined with the location-based dissemination procedure.

4.6.1 Simulation setup

The simulation environment is based on the microscopic road traffic simulator SUMO coupled with the event-based network simulator NETSIM as described in the previous chapter. An area of $2 \times 2km^2$ of an European-like city center (namely, Toulouse, France) using Open Street Maps (OSM) is considered. It exhibits an irregular road structure and the presence of large buildings affecting the quality of wireless transmissions. The vehicle density is varied between 30 to 500 vehicles. The maximum transmission range of each vehicle is set to $R_{max} = 250m$. The number of RSUs is varied from 2 to 8, as well as their position. Our algorithm is implemented using the Python language and is run on an Intel Corei5 2GHz and 8GB RAM system. The convergence time of all simulations is below 0.7s. Table 4.2 lists the parameters used in the evaluation.
Finally, gray zones in the considered map are identified by simulation as follows. First, RSUs are configured to broadcast alert messages every $100ms$ for $500s$. Next, the average Message Delivery Ratio (MDR) is computed for each square. As road safety applications require high reliability between 90% and 95% [Boban 2017], squares with an MDR below 90% are considered gray.

Table 4.2: Simulation configuration parameters

Parameter	**Value**
Discount rate γ	0.9
Learning rate α	0.1
Epsilon ε	0.1
Simulation time	500 s
AM generation start	10 s
AM generation rate	10 packets/s
AM packet size	1024 bytes
Propagation model	Nakagami m = 3
d_{max}	16 m

Figure 4.3: Initial environment matrix

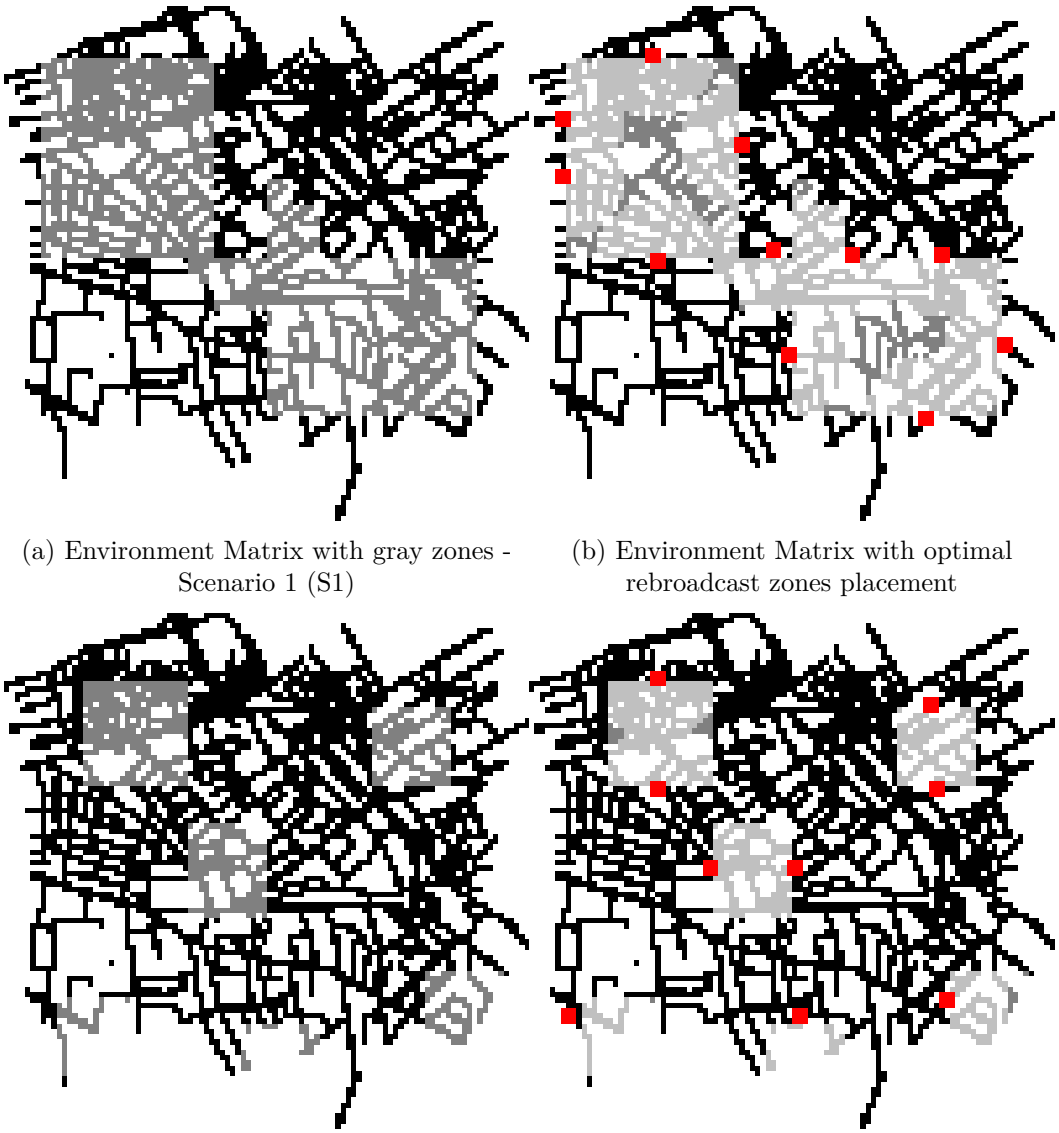
We start by converting the geographic map in Figure 4.2 into a matrix of 50×50 of $40 \times 40m^2$ squares, with values "0" (road zone could be elected as rebroadcast zone) and "1" (non-road zone or zone X) for each square as mentioned in Figure 4.3. Then, we insert the gray zones resulting from the simulations explained above in the concerned road zones (switch values from "0" to "0.5" (gray zone)), as presented in Figure 4.4a. Next, we implement and apply our environment's Q-learning algorithm described above. The initial number of rebroadcast zones "N" is calculated as the ratio of gray squares and the average number of gray squares that a rebroadcast zone can cover (6×6 squares: approximately the transmission range of a vehicle in a rebroadcast zone center).

4.6.2 Performance evaluation of our Q-learning placement method

Three scenarios are considered with 2, 4, and 8 RSUs to address three different situations: an insufficient number of RSUs to decently cover the whole area (with numerous large gray zones), a decent number of RSUs, and finally, a high number of RSUs leading to multiple scattered small gray zones.

Figure 4.4b and 4.4d shows the positions of the "N" optimal rebroadcast zones (red squares: we have enlarged the size of the rebroadcast zones in the figure for visibility) for scenario 1 and 2; after the convergence of the algorithm and the gray squares (light gray squares) covered by these latter.

The first scenario (S1) corresponds to the simulation settings in the previous chapter with 2 RSUs. Our RL algorithm computes 11 rebroadcast points covering $\approx 84\%$ of gray zones, while with the default method, 13 are obtained with a coverage of $\approx 80\%$. Full coverage is not achieved as 2 RSUs are insufficient to reach all gray zones, assuming one rebroadcast from a relay vehicle.



(a) Environment Matrix with gray zones - Scenario 1 (S1)

(b) Environment Matrix with optimal rebroadcast zones placement

(c) Environment Matrix with gray zones - Scenario 2 (S2)

(d) Environment Matrix with optimal rebroadcast zones placement

Figure 4.4: Rebroadcast points placement with our Q-learning method

(a) Initial RSU placement (Scenario3) (b) With the default method (c) With our Q-learning method

Figure 4.5: Rebroadcast points placement

The second scenario (S2) considers 4 RSUs placed at different locations. Our Q-learning algorithm leads to the full coverage of gray zones with only 9 rebroadcast zones, while 17 are needed with the default one.

As illustrated in Figure 4.5a, the last scenario (S3) considers 8 RSUs. The transmission power (i.e., transmission range) of RSUs and vehicles is reduced. With the default method, we have 48 rebroadcast zones (Figure 4.5b) vs. 15 (Figure 4.5c) with our algorithm, which achieves a better coverage than the default placement, around $\approx 93\%$ of gray squares.

4.6.3 Performance evaluation

This section details the performance results for Toulouse scenarios (S1 and S2) under various vehicle densities using the LAMD dissemination procedure, comparing both techniques (the default placement and the Q-learning proposed placement).

4.6.3.1 Performance metrics

- **Information Coverage (IC):** computed as the total number of vehicles that successfully receive an AM at the end of the simulation (NumVehRecMsg) divided by the number of vehicles (TotalNumVeh) averaged on all generated AMs. This metric shows how successful the dissemination is after a decent period of time.

$$IC = \frac{NumVehRecMsg}{TotalNumVeh} \tag{4.4}$$

- **Dissemination delay:** The dissemination delay is the total time required to deliver the AM to all the vehicles in the area of interest that receive the AMs. The vehicles that do not receive any AM are excluded from the computation. This metric measures how fast the dissemination can reach the vehicles within

the area of interest.

- **Collision Ratio (CR):** The collision ratio is the percentage of MAC collisions (NumCollisionPkts) divided by the number of packets sent (NumSentPkts) computed over the simulation duration.

$$CR = \frac{NumCollisionPkts}{NumSentPkts} \tag{4.5}$$

- **Redundancy Rate (RR):** The average number of AM rebroadcasts or duplicate transmissions (i.e., NumDuplicPkts) out of all sourced (generated) AM (NumSPkts) [Panichpapiboon 2011] .

$$RR = \frac{NumDuplicPkts}{NumSPkts} \tag{4.6}$$

4.6.3.2 Performance analysis

Information Coverage: Figure 4.6 presents the Information Coverage as a function of vehicle density for the LAMD technique in both scenarios (S1 and S2), comparing the default and RL-based rebroadcast zones placement. For scenario n°1, results are almost similar, with a slight shift of 0.2 to 0.5% between the default and the new RL methods. However, for scenario n°2, we notice that the RL method is significantly better than the default method with low traffic density due to the accurate and optimal placement of rebroadcast zones; as cited above, only 9 rebroadcast zones are needed with the RL algorithm vs. 17 with the default one, despite requiring approximately half the number of zones. When the density increases, we achieve maximum coverage in both methods. The results show that the RL method ensures an accurate placement with a minimum number of rebroadcast zones required to achieve full coverage by applying the LAMD. However, with the default method, we also have good results. Still, it remains a theoretical method with an inaccurate placement and an approximate number of rebroadcast zones, which could be much by increasing the number of RSUs.

Dissemination Delay: This effective placement also affects the dissemination delays, as shown in Figure 4.7. For low vehicle densities, where the rebroadcast zone placement particularly matters, the difference between the dissemination delays of the two methods is at least $100ms$, meaning that at least one additional rebroadcast (i.e., of the next instance of an AM) is needed with the default method compared to our method. When the traffic density increases, the probability of the presence of vehicles in or around the rebroadcast zones increases. As the RL-based placement ensures the full coverage of gray zones, more vehicles are reached from the first AM rebroadcast. Indeed, starting from a vehicle density of 100, with the RL-based placement, on average, all vehicles are reached within $100ms$. In comparison, more than twice this density is needed to achieve such performance with the default method.

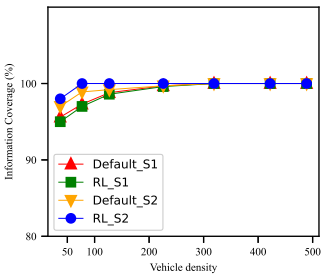


Figure 4.6: Information coverage

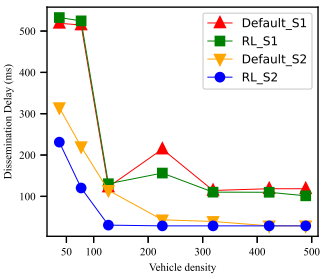


Figure 4.7: Dissemination delay

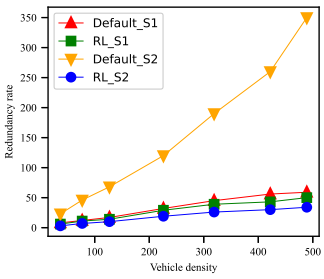


Figure 4.8: Redundancy ratio

Redundancy Ratio: Figure 4.8 shows the number of duplicated AM transmissions versus vehicle density for LAMD technique in both scenarios (S1 and S2), comparing the default and RL-based rebroadcast zones placement. As expected, with the RL-based placement, the redundancy is significantly limited, by choosing the minimum number and the optimal location of the rebroadcast zones to serve the gray zones, especially in Scenario n°2, where theoretically, we cover 99% of the rebroadcast zones (as explained above) with just 9 rebroadcast zones. However, by increasing the number of RSUs in scenario 2, rebroadcast zones increase respecting the default method, which increases the number of relay vehicles in the different rebroadcast zones, hence the redundant number of messages (Several vehicles can receive the same message more than once from different relay vehicles). Indeed, fewer relay vehicles are eligible for an AM rebroadcast by minimizing the number of rebroadcast zones and optimally choosing their locations to serve the gray zones. This avoids useless redundant rebroadcasts.

Whatever the placement method, in all the considered scenarios, the collision ratio remains very low ($< 0.03\%$ vs. $< 0.0001\%$ with the RL-based method), even for high vehicle densities. This is mainly due to the AM dissemination procedure, which drastically limits the contention when rebroadcasting an AM. Indeed, the relay vehicle selection limits the set of vehicles that can act as a relay to those in close vicinity to rebroadcast points. In addition, it further manages contention between nearby eligible relay vehicles by assigning different back-off waiting periods before pursuing a rebroadcast attempt.

4.7 Conclusion

This chapter has introduced a new Q-learning-based method that provides the location-based AM dissemination procedure with the minimum number and optimal locations (rebroadcast zones) where vehicles are invited to rebroadcast an AM in order to deliver it on a pre-defined region, which may include multiple gray zones. As a result, our method provides the best possible AM coverage, fast AM delivery, and very limited redundant and useless AM re-transmissions (i.e., network overhead). Our simulations assess the performance gains of our placement

method on a real portion of a European city center with realistic road traffic models. However, we can apply our algorithm to larger maps (e.g., in the case of cellular networks, where the coverage of the Base Station (BS) controller is more important), as the complexity remains practically the same.

In this chapter, the gray zones were inputs for our algorithm. In the next chapter, we will propose a new method based on machine learning algorithms to approximately locate these zones by estimating the quality of links in each square, considering several factors.

Chapter 5

SDN-enabled Machine learning Road Link Quality Estimation

5.1 Introduction

In order to support ITS services with strict QoS requirements, despite rapidly changing traffic conditions, the network must be able to anticipate possible changes in the performance of some of its links (changes related to the evolution of traffic, such as the increase in the density of nodes in a given area, etc.) and take the required actions, to optimize its operation and performance and to work towards guaranteeing continuity of services with the required QoS. Hence, we speak of anticipatory/proactive control of the network [Bui 2017].
Such control requires an estimated view of the network's future state, allowing one to consider and realize a proactive and intelligent network control.

In this chapter, we propose a Infrastructure to Vehicle (I2V) link quality estimation technique on the road, which constitutes the first essential step for our alert message dissemination scheme's efficiency, specifically for rebroadcast zone placement. The latter relies on Machine Learning techniques to estimate the link quality (Packet Reception Rate (PRR)) in each zone (defined by a square as described in the previous chapter), which allows us to determine the gray areas to place our relay zones. A supervised learning model has been proposed, with the particularity of using traffic information and "Hello" default exchange messages between vehicles and infrastructure, thus limiting transmissions' overhead.

In order to train and evaluate the proposed model, a dataset was generated, as the ones available online do not fit the needs of our study. This dataset was generated using mainly the NETSIM framework and based on mobility traces.

This chapter is organized as follows. Section 5.2 presents general motivations for link quality estimation problem. Then, we present a synthesis of existing works in the scientific literature in Section 5.3. Section 5.4 gives a general overview of the proposed link quality estimation model. Section 5.5 details the proposed models, while the next section describes the dataset used. Section 5.7 focuses on the experimental part. It first presents the metrics considered, then analyzes the evaluation results for the proposed model. Finally, the last section concludes this chapter.

5.2 Link quality estimation: motivations

In vehicular networks, the conditions of the radio signal propagation channel can vary considerably in time and space, which affects the link quality on the road. In order to ensure reliable and durable performance in such networks, efficient estimation of the link quality on the road is required by some dissemination techniques so that the parameters of the link and the choice of relays can be adapted and an alternative or more reliable route/area can be selected for data retransmission. In other words, the better the link quality is, the higher the successful reception rate and the more reliable the communication are. However, challenging factors that directly affect the quality of a link, such as channel variations, send/receive problems, and complex interference patterns, to name a few, can inevitably lead to unreliable links [Cerar 2021].

On the one hand, it is difficult to incorporate these changing factors into an analytical model. Therefore, such models cannot be easily adopted in realistic networks due to the design environment's highly arbitrary and dynamic nature. On the other hand, effective link quality prediction can provide significant performance, such as improving network throughput due to reduced packet loss, extending network lifetime due to limiting retransmissions, limiting topology outages, improving reliability, etc. Finally, variations in the link quality can significantly influence the overall connectivity of the network. Therefore, the efficient estimation or prediction of the link quality can provide the best-performing link from a set of candidates to be used for data transmission.

5.3 Related work

During the last decade, wireless link quality estimation and characterization have been studied in wireless mobile networks and considered crucial for reliable communication [Tang 2007, Cerar 2021]. Due to the dynamic nature of the vehicular environment, the statistical channel models do not predict wireless link quality with high accuracy [Tang 2007]. Therefore, advanced methods' link quality prediction has become popular among researchers. Due to the increasing computational ability of network nodes and the data handling capacity of clouds, artificial intelligence algorithms have found new application domains. The term link quality can be generally thought of as a link description derivative of throughput or reliability. Unfortunately, there is no standard definition for link quality nor a standard unit of measure for the metric [Baccour 2012]. Depending on the context, it can be a qualitative or quantitative link description.

In most cases, link quality is expressed in a probabilistic sense of the link's reliability in the past or expected packet delivery. Naturally, the measure of such a probability would range from 0 to 1. In other cases, link quality may be described as a percentage of maximum throughput potential or packet reception and may vary from 0 to 100. And finally, link conditions are often categorically described using natural languages such as 'good,' 'intermediate,' or 'bad.' The link quality measure

depends on the application and its intended use [Lowrance 2017].

Some comprehensive surveys have been conducted to prove the potential of machine learning techniques in solving various problems and predicting link quality in wireless communications [Mao 2018, Cerar 2021].

Most link quality estimation techniques in vehicular networks are proposed to reactively estimate the quality of the V2V links in order to select the next hop/broadcaster between the sender and its neighboring nodes [Bauza 2013, Okada 2009, Cai 2014, Wang 2012], where the received signal's strength or packet reception rates over a given node's link have been used to characterize the quality of its forward link. However, such mechanisms have assumed a fixed communication range among the nodes, which is not realistic [Amoroso 2011, Wang 2015]. In addition, the rebroadcaster links' qualities can considerably vary for a given node in time for several reasons, such as varying surrounding node densities and fading channel effects [Rak 2013].

Machine learning (ML) based techniques have recently emerged as an alternative to link quality estimation techniques based on predefined models [Ye 2018]. This avoids simplifying assumptions. In [Turan 2021, Ramya 2019], authors propose ML-based algorithms to predict V2V path loss, proving that such models' application offers better performance than traditional analytical models utilization (e.g., log distance path loss model). Benrhaim and al. [Benrhaiem 2020], also focus on V2V links. Their method relies on beacons periodically generated message exchanges between vehicles to estimate V2V link reception quality. They propose a Bayesian network-based scheme at different locations in the zone covered by the transmission range of the sender for the estimation. They model the network as a grid-like city streets plan and suppose that the length of a road segment and the transmission range of vehicles are the same. First, a training data collection phase is executed by regular vehicles and a coordinator (a vehicle located around the center of a road segment) to collect instantaneous reception status (whether beacon reception is successful or not) information in the transmission range. Next, each coordinator processes the received beacons of the vehicles in the current road segment and then executes a Graphical Model Learning (GML) to estimate V2V link reception quality in the transmission range (road segment) using Bayesian networks. A Bayesian network is a directed, acyclic graph that discovers and defines relationships among random variables from observational data for prediction. This is the only work that estimates the road links' quality. Estimation results show good accuracy. However, they only consider the reception messages status exchanged between vehicles and used the same parameters in all simulations for data collection. But in reality, other parameters can influence the road link's quality, such as traffic density, the distance between nodes, etc. In addition, the estimation is done in a distributed way by a road coordinator (vehicle in the center), which is unreliable with the mobility of the vehicles; the position changes rapidly.

Most of the works proposed in the literature to estimate the vehicular communication links' quality assume simplifications of vehicle mobility. We note these works concern V2V links, where only one is interested in estimating V2V link

quality on the road, and approaches based on machine learning techniques generally present the best performances for both problems.

Our work is the first one that focuses on I2V wireless quality links on the road. The proposed method for estimating road links' quality excludes any vehicle mobility or communication range assumption. Instead, each entity (SDN controller) runs its machine learning model trained using the road traffic data of the area covered by that entity. The proposed approach is detailed in the following sections.

5.4 Proposed model overview

In our work, we propose a model based on machine learning techniques M_{PRR}, allowing us to estimate the quality of Infrastructure to Vehicles (I2V) links on the road, precisely, the Packet Reception Ratio (PRR) of RSUs' transmission for each zone in a predefined region. This zone is defined by a small square, as described in the previous chapter. As described in Chapter 3, these squares are used to identify gray zones in the region where their PRR is less than a predefined PRR threshold (90% in our case) [Meireles 2010]. The model mainly combines the information reported by the RSUs and the vehicles during the exchange of periodic beacon messages (or "Hello" messages) between the two entities and road traffic conditions (Vehicles density on the road). Furthermore, this model is executed by each SDN controller, which manages a collection of nearby RSUs, allowing it to learn the link quality variations in its coverage area. Finally, this information is combined with other data to extract features used as input to the model to estimate PRR in each small zone. The design of the model is detailed in the next section.

Figure 5.1 shows the key elements of the proposed approach. First, we assume all vehicles are equipped with a GPS module and can send information, such as their position ($P(x, y)$) and the packet response of the "Hello" message received from the RSU at the time of association and beacon messages exchanges. This information is periodically collected by each RSU and hosted in the cloud or shared directly with the network controller to extract features used as input to the model in order to estimate the PRR. Finally, we assume that an RSU entity reports information about each newly associated vehicle to the SDN controller.

5.5 SDN-enabled machine learning road I2V link quality prediction

This section briefly introduces the main principles of Machine learning (ML) techniques and then details the proposed model's design.

5.5.1 Machine learning techniques

In 1959, Arthur Samuel, a Machine Learning (ML) pioneer, defined ML as "the field of study that gives computers the ability to learn without being explicitly

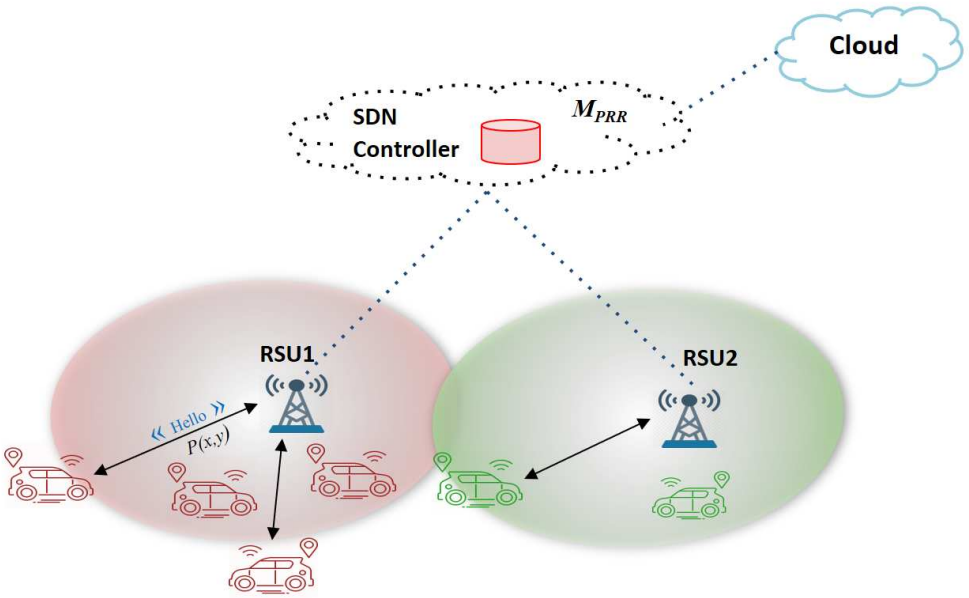


Figure 5.1: Key elements of the proposed approach

programmed." From this definition, we understand that an ML system, unlike a computer program, is not programmed with a set of instructions to execute based on inputs. But instead, a system learns the best actions to take (decision, prediction, etc.), usually from a set of data or past experiences.

These systems are classified according to several criteria. Among these criteria, we cite the nature and importance of the supervision provided to the system during the training process. We distinguish three main categories: supervised learning, unsupervised learning, and reinforcement learning (described in the previous chapter) [Kotsiantis 2007].

- *Supervised learning*: learning is performed from a labeled dataset, so the training data contain the desired solutions. An example of a supervised system is classification; consider a spam detection system. The training data contain many examples of emails with a label for each mail, indicating whether or not it is spam. After the training phase, the system can predict each new mail's corresponding label (spam or not).

- *Unsupervised learning*: unlike the supervised type, the dataset is not labeled. For example, a dataset listing the information of the customers of an e-commerce site (age, gender, time of visit, category of items seen and/or purchased, etc.). The system's objective is to identify similarities between the various customers and to group them into clusters, for example, for advertising purposes (personalized offers/discounts, etc.). These techniques are known as "clustering" in the literature.

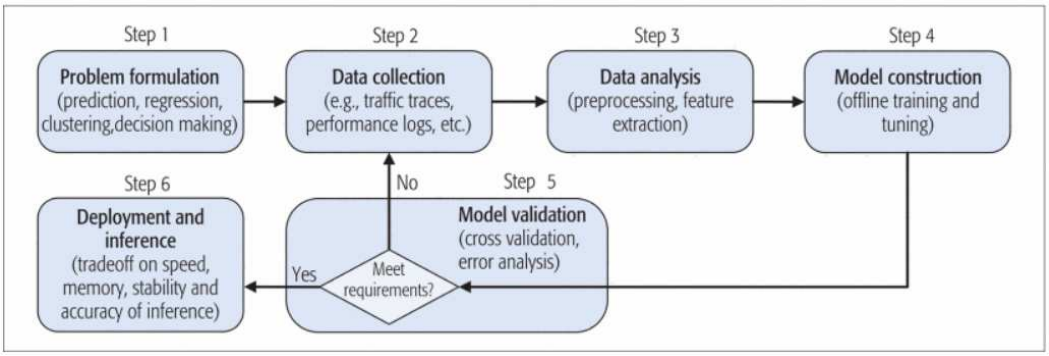


Figure 5.2: The typical workflow of machine learning for networking [Wang 2017]

- *Reinforcement learning*: learning is run from previous experiences, not from a data set. We talk about an agent that evolves in an environment and makes decisions. In return, it gets a reward (or a penalty, as a negative reward). The model aims to find the best policy to maximize the cumulative excepted return. An example is a robot that tries to reach a given destination while avoiding obstacles on its way.

In our study, we consider supervised learning, where training is performed using a labeled dataset. Formally, in a dataset D defined by $D(x_1, y_1), ...(x_n, y_n)$, the training of the model M aims to find the best relationship between the inputs X, called predictors, and the outputs y, called label, $y = M(X)$, such that, for new input data X_n whose outputs are unknown, the model can predict the corresponding output $\widehat{y_n} = M(X_n)$ with good accuracy.

We distinguish two types of supervised learning: regression, when the value to be predicted is a continuous real number, $y \subset R$ and classification, when y belongs to a finite set $C = c_1, c_2, ..., c_n$ called classes. In the spam detection example (presented earlier), two classes are considered (spam and normal).

Several supervised learning techniques have been proposed in the literature. Some are specific to regression (i.e., polynomial regression), others are dedicated to classification (e.g., K-Nearest Neighbor K-NN), and others are adapted for both categories (e.g., decision tree). Each technique has advantages and disadvantages (type of data processed, sensitivity to noise in the data (e.g., anomalies), training time, resource consumption, etc.). A detailed comparative study is presented in [Kotsiantis 2007, Klaine 2017].

The typical ML application to networking workflow is presented in Figure 5.2 specified in this work [Wang 2017].

In addition, we consider the so-called ensemble learning techniques [Sagi 2018], allowing us to train several models (of the same or different techniques) and combine their predictions. This technique represents one of the most popular and powerful supervised algorithms. It allows for the design of a generalized model and avoids overfitting (when the model has understood the problem taught during the training phases, but it is based only on this initial training data. It then suffers from a

generalization problem and cannot recognize other data). The model proposed in our studies is based on this technique. In particular, we use the *Random Forest* technique [Breiman 2001], which trains a set of decision tree models. This technique is used for both regression and classification problems. Despite its interpretability (i.e., it has a simple structure to explain (in terms of relations between inputs and outputs), compared to a neural network, for example), it offers several advantages that motivated our choice. First, it allows dealing with problems with several classes, compared to other techniques focusing only on binary classification (i.e., two classes). This is called multivariate classification. Furthermore, it integrates a feature to select the best predictors to consider, called "feature importance." From the constructed trees (a process explained in detail later), the most important features are likely to appear in the nodes near the tree's root, while the others are often closer to the leaves (or do not appear in the tree).
Moreover, it is less sensitive to outliers and is fast in execution. Indeed, its complexity depends on the depth of the trees used. But it can handle non-linear data with a large number of features. Besides, it is less sensitive to data overfitting and exposes some parameters limiting the growth of the trees (detailed later).

As presented above, our problem concerns the Packet Reception Ratio (PRR) estimation. This problem has been modeled as a regression problem. The estimated variable is the PRR in each map's little zone (square) under the SDN controller coverage. In the following sections, we present the considered training variables (or features) to design our models. These features are designed according to the objectives, i.e., i) features that require a minimum of information from the vehicles and ii) features that are independent of the network technology used. Next, we detail the techniques used to train and calibrate the model parameters.

5.5.2 Proposed framework

The propagation channel characteristics of vehicular networks differ significantly from other wireless systems. The physical environment in vehicular channels is expected to experience random variations caused by several factors, including mobility patterns and rapid changes in traffic density, path loss effects, and environmental effects. The rapid temporal variability and non-stationary channels are rationales for developing a unique framework for predicting the link quality of vehicular networks. In this work, we aim to build a machine learning model capable of predicting the PRR on the road with minimum error. Machine learning models are ideally suitable for classification and pattern recognition problems.

By following the typical Machine learning workflow specified in [Wang 2017] and leveraging SDN [Feamster 2014, Foster 2020], we come up with the SDN-enabled machine learning PRR prediction framework in Figure 5.3. The machine learning workflow includes six stages: Problem formulation, Data Collection, Data Analysis, Model Construction, and Model Validation, and the last stage is Deployment and Inference.

The framework is based on two main pillars: SDN and the power of suitable

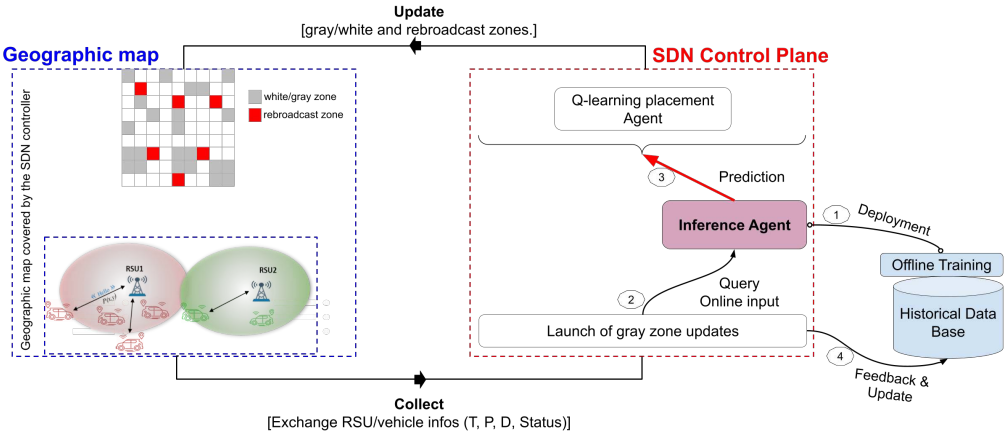


Figure 5.3: Packet Reception Ratio (PRR) inference framework

machine learning algorithms to learn some properties of an historical dataset and leverage the learned proprieties to provide good estimations on new observations. From Figure 5.3, the framework's workflow is as follows. Firstly, the prediction model is constructed offline by training and parameter tuning the historical data. The historical dataset may be composed of a large number of samples. Each sample represents a combination of features' values and the associated target value since we are in a supervised learning configuration. Collected data include the transmission power of RSUs (T), vehicle position (P) (to identify the zone identifier (Z_i) and calculate the distance to the RSU), the distance between the vehicle and RSU (D) (to identify the average distance between the concerned zone and the RSU), and packet status ($Status$) to calculate the target by zone (whether or not the vehicle has received the "Hello" messages from RSU). The features' description, collection, and processing will be detailed in the next section.

The designed model is then deployed (1) as the Inference Agent. The model is deployed to be used for PRR inference. The PRR is used to predict gray zones in the region to update the rebroadcast zones for disseminating alert messages according to traffic needs and conditions, as described in the previous chapter. We can suppose that this approach can be used proactively, i.e., we assume the controller has a historical idea about traffic conditions change (e.g., traffic density at peak hours), so it updates these zones according to these schedules (e.g., every two hours or three times a day).

The online input for each zone after processing (2), composed of $(Z_i, T, P_{loss}, V_d, D)$ (where V_d is the vehicle density in the concerned zone Z_i, calculated according to the real-time number of vehicles in the zone/square, and P_{loss} is the packet loss in the zone calculated with respect to real-time parameters and traffic conditions), is got when the controller launches the updates of gray zones for the geographic map under its coverage. Taking this input, an inference of the PRR in each zone (small square) is made, so gray zones are identified (3). The information will then be used as input by the Q-learning placement Agent to

update rebroadcast zones as described in the previous chapter.

Finally, when the process is complete, its observed output is also collected efficiently, and the historical dataset can be updated with the newly collected data (4). Therefore, having the database up-to-date is essential and will allow considering new dynamics from the traffic changes. The historical data gathering and online update of the historical data with the newly collected data form a base for our framework. The historical data could be enriched from the cloud.

5.6 Learning-based modeling

According to the workflow from [Wang 2017] this section begins with the problem formulation. For the PRR inference, the target metric is a continuous variable; its prediction is a regression problem.

5.6.1 Dataset and analysis

As indicated above, our model based on supervised learning requires a training phase using a dataset. In our study, we focus on the case of VANET in an urban environment. In order to evaluate the proposed model in a realistic framework, it is necessary to have a data set that meets the following requirements:

- Network coverage: Consider cells with varied communication ranges (small and large coverage) by manipulating the transmission power of RSUs. In addition, it is necessary to know the geographical positions of these entities to calculate the distance.
- Road traffic: Have vehicle data (location) driving on most roads (main and secondary) covered by a given RSU.
- Size of the dataset: A collection of data for a long duration by varying the density of the vehicles, the transmission powers, the position of the RSUs, and the coefficients of path loss, allowing to explore the temporal variations of the measured metrics.

In vehicular networks, various data collection campaigns have been realized by the scientific community in order to produce usable data sets (e.g., to study the performance of these networks). However, very few of these data sets are available in open access. Moreover, they do not consider urban mobility and do not have all the parameters required by our model.

Since no dataset carried out by the scientific community fully corresponds to our prerequisites, we generated our dataset. However, technical constraints and time limit the realization of a data collection campaign. Therefore, we move towards an approach based on simulation tools/mobility emulation and VANET networks.
The dataset used in this work was generated using the microscopic road traffic simulator SUMO [Busanelli 2013] coupled with the event-based network simulator

NETSIM as described in [Weber 2021]. NETSIM simulates the DSRC protocol stack (signal strength, handover, connectivity), while SUMO handles vehicle mobility. The global framework provides a realistic simulation of DSRC connectivity for vehicles. Our simulation setup comprises two main parts: the first concerns implementing the DSRC network, and the second concerns the implementation of vehicle mobility. In this work, we extract the area of $2 \times 2km^2$ of a European-like city (namely, Toulouse, France) using Open Street Maps (OSM). The choice of the area is significant as it is present in the center of the city with high traffic densities (Urban environment), large buildings affecting the quality of signals on the road, and irregular road structures. The random Trip application of the SUMO package is used to auto-generate the trips for the vehicles with mobility over the given map area. We assume that all vehicles are equipped with wireless communication modules DSRC. Regarding the DSRC network, we place 4 RSUs in the selected region, as shown in the Figure 5.4.

A wireless transmission's success depends on multiple factors: distance, transmitter power, path loss, fading, receiver sensitivity, etc. Moreover, depending on the environment (Highway, Urban, Obstructions, Line of Sigh), the transmission coverage of an RSU may vary widely, for example, from 100 to 700 meters, for the same transmission power shown in [Gozálvez 2012, Böhm 2010]. For network simulation parameters, we run 196 simulations where each RSU broadcasts a control message every 100 ms for 500s (27 hours in total), varying each time the transmission power, the positions of the RSUs, vehicle densities, and the path loss coefficients. Table 5.1 lists different manipulated parameter values. Path loss is the reduction in the power density of an electromagnetic wave as it propagates through space. It may be due to many effects, such as free reflection, aperture-medium coupling loss, and absorption. The path loss exponent varies between 2–5 depending on the surrounding environment coherence bandwidth and Doppler shift. In reality, the path loss exponent can be calculated according to several parameters (e.g., the transmission power, the distance from the sender, the wavelength of the carrier wave (which depends on whether and obstacles), etc.), but as we work with data collected by a simulator, we use this parameter directly from the simulator. However, an SDN controller can use its global view and knowledge to calculate this parameter based on the real-time information (distance between RSUs and zones, urban or other environments, weather, etc.) shared by the RSUs and the cloud [Zakaria 2021]. After each packet, we record the vehicle position (to identify the zone identifier), the vehicles that received the message, and the packet status (success or error). The results of simulations generate a dataset of 52007 observations. Figure 5.5 shows the number of samples generated for each PRR range. The data set generated and collected after each simulation are described in Table 5.2.

We represent the geographical map as a 50×50 grid matrix of $40 \times 40m^2$ (Figure 5.4). Thus, we predict the PRR by a small road zone/square (only squares containing road areas are considered) instead of a road because the quality of links at the beginning of the road could differ from the end and the center. Hence, predicting the PRR by zone (a small portion of the road) is more precise and

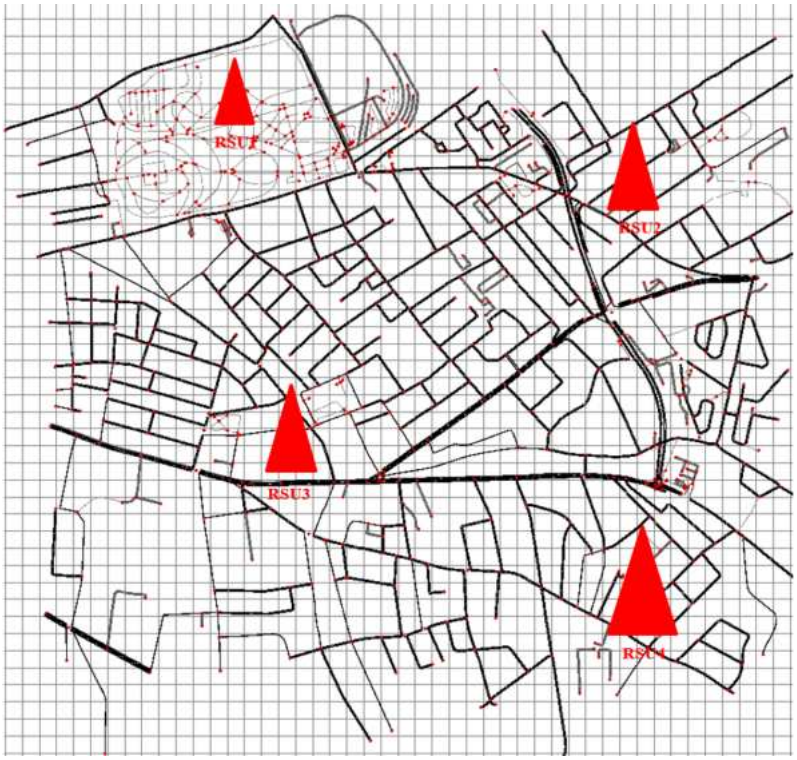


Figure 5.4: Map with geographic locations of RSUs

Table 5.1: Simulation configuration parameters

Parameter	Value
Vehicle Density	from 5 to 500 vehicles/zone(square)/hour
Path loss coefficient	from 2 to 5
Distance	from 0 to 2 km
Transmission power	from 10 to 50 dbm

accurate. We realize some data processing to have our dataset suitable for our model, which allows us to predict the PRR in each zone as described in Table 5.3. First, we consider the vehicle's position to define the concerned zone. Then for each simulation, we calculate the density of vehicles per zone and hour, the average distance between the zone and the sender, and the PRR (the number of packets received by vehicles in the zone compared to the number sent by RSUs to those vehicles).

5.6.2 Model training

In the previous section, we presented the different learning variables used by our model M_{PRR}. M_{PRR} mainly uses vehicle density, distance, transmission power, and path loss exponent features. These features are designed according to two main criteria i) features requiring a minimum of information from the vehicles and

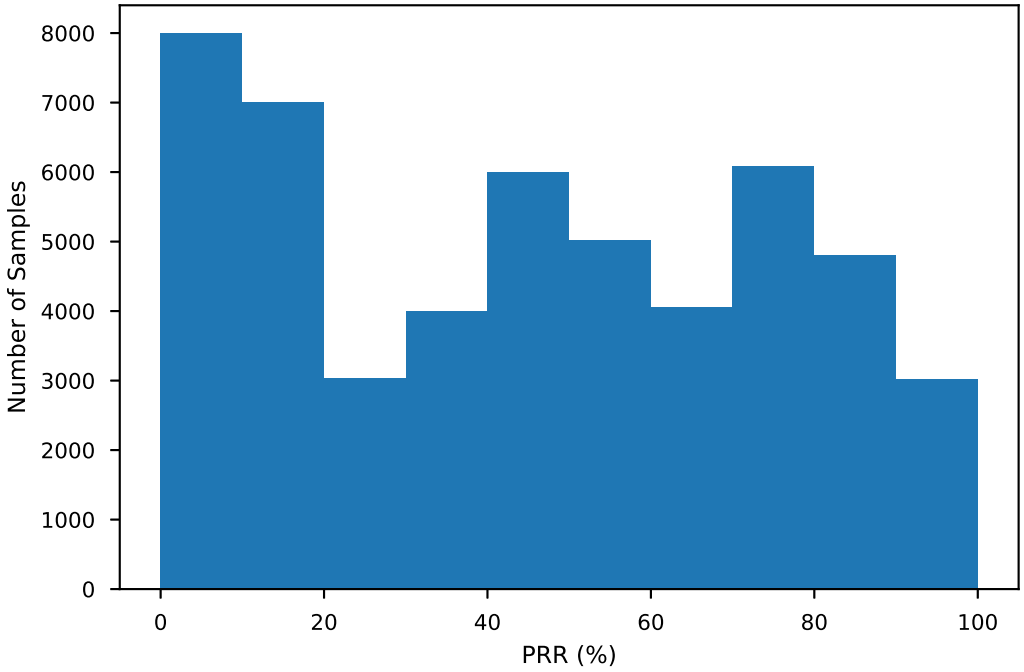


Figure 5.5: Number of samples generated for each PRR range

Algorithm 3: Road PRR Estimation

Input:

$Z_i, i \in 0, .., N$ List of road zone ID

$P_{loss}(i)$ per *zone*: path loss per zone

$T(i)$: Transmission power of the RSU that covers the zone

$V_d(i)$: Vehicle density in the concerned zone

Output:

$\widehat{PRR}(Z_i)$: the Packet reception ratio by zone

1 **for** $i = 0$ *to* $i = N$ **do**

2 $D(i) = \sqrt{((x_{RSU} - x_i)^2 - (y_{RSU} - y_i)^2)}$ `/* distance between the zone` i `of coordinates` (x_i, y_i) `and the RSU that covers it of coordinates` (x_{RSU}, y_{RSU}) `*/`

3 $\widehat{PRR}(Z_i) = M_{PRR}\ (D(i),\ P_{loss}(i),\ T(i)\ ,\ V_d(i)\ )$

Table 5.2: Generated and collected data set

Feature Name	**Feature Description**
Packet Id	sent packet identifier
RSU ID	sender RSU identifier
Vehicle ID	receiver vehicle identifier
Vehicle position	coordinates (x,y), which can be converted a GPS coordinates
Transmission power	Transmission power of the sender RSU
Path loss exponent	Power density of an electromagnetic wave
Packet status	Packet Status "received" or "not-received"

Table 5.3: Parameters and notations

Feature Name	**Feature Description**
Zone Id	Zone identifier
Vehicle density	the density of vehicles per hour in the concerned zone
Distance to the sender RSU (m)	Average distance between the sender and the zone
Transmission power	Transmission power of the sender RSU
Path loss exponent	Power density of an electromagnetic wave
PRR	Packet Reception ratio in the concerned zone

ii) independent of the network technology used.

$$\widehat{PRR(Z_i)} = M_{PRR}(V_d(i), D(i), T, P_{loss}(i)) \tag{5.1}$$

where,
Z_i : Zone identifier or zone position
$V_d(i)$: The density of vehicles in the concerned zone
$D(i)$: Distance between the RSU covering the zone, and the zone
T : Transmission power of the RSU
$P_{loss}(i)$: Path loss in the concerned zone

Algorithm 3 summarizes the information used as input and the features considered by the M_{PRR} model to estimate the packet reception ratio for each road zone covered by the SDN controller. Note that the zone ID list and distance from the RSU are already saved in the controller database as they have fixed positions.

From a dataset composed of the various attributes listed below, labeled by the

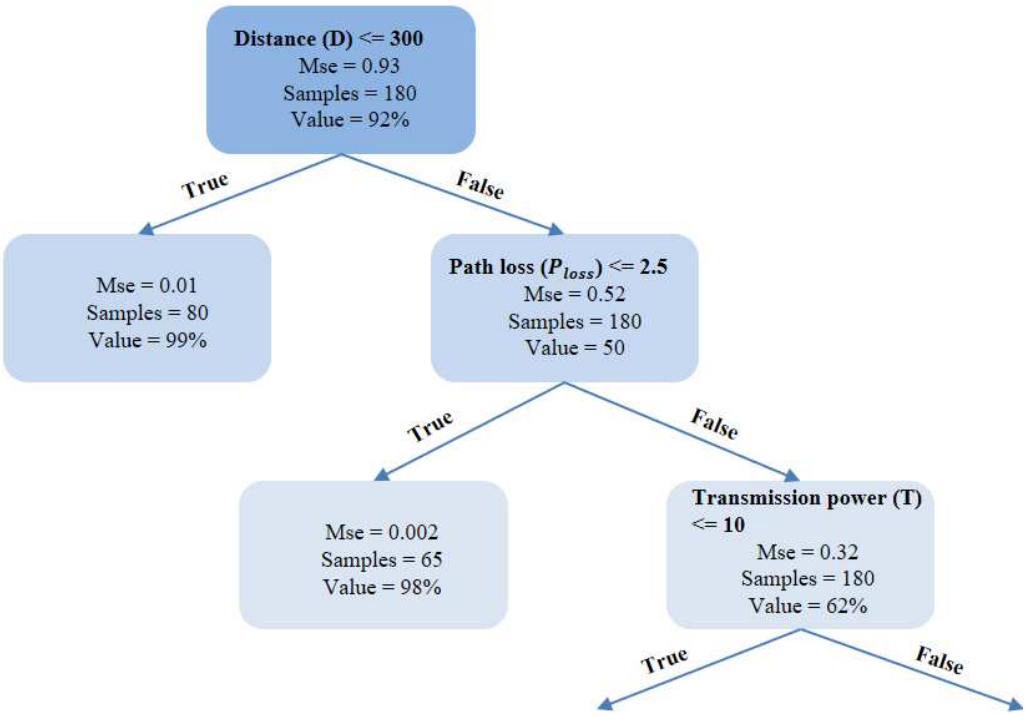


Figure 5.6: Simplified example of decision trees (D.T)

desired results (PRR), a training of the models is performed offline to find the best representation between features and labels. As mentioned above, our models is based on the Random Forest (R.F) technique. This technique combines several Decision Tree models (D.T). In a decision tree, data are structured as a tree from which the model makes predictions for new data. Then, according to the input data, the predictions are made by traversing the tree from its root to a terminal node (i.e., node without a leaf), called a leaf. These contain the values of the predictions (packet reception ratio PRR of the model M_{PRR}. In our case (a regression case), the average of the node observation values is used as a prediction. Figure 5.6 represents an example of a decision tree [1]. The first condition of the root node is verified. It concerns the distance from the RSU (D); if it is less than or equal to 300, the predicted value of the PRR is the average of the observations of the left child node (i.e., a terminal node), which is 99% (average calculated with 80 observations). Otherwise, the right child node is traversed if the distance D is strictly greater than 300 meters. Since it is not a terminal node, a new condition is checked. This time, it checks if the path loss exponent is less than 2.5. If this case, the predicted value is 98% (terminal node on the left); otherwise, the child node on the right is crossed to check new conditions, and so on.

The trees' construction is the basis of the predictions. This is the main task of

[1]This is a simplified example. The values are provided for explanation purposes and do not represent the data used by our model.

the training phase. During this phase, the model defines the nodes, the number of observations (samples) per node, and the rules to check for each node. For each node, the model looks for the pair (k,t_k) (with k the attribute to consider (distance, density, transmitter power, etc.) and t_k the value of this attribute) that minimizes the mean square error (MSE) using the cost functions represented by equation 5.2.

$$j(k, t_k) = \frac{m_{left}}{m} MSE_{left} + \frac{m_{right}}{m} MSE_{right} \tag{5.2}$$

Where,

$$\begin{cases} MSE_{node} = \sum_{i \in node} (\hat{y}_{node} - y(i))^2 \\ \\ \hat{y}_{node} = \frac{1}{m_{node}} \sum_{i \in node} y^{(i)} \end{cases} \tag{5.3}$$

$p^2_{i,k}$: represents the percentage of observations of class k among all training observations in the i^{th} node,

$m_{left/right}$: represents the number of instances of the left/right subset.

Our Random Forest model trains a set of decision tree models. First, each tree is trained with a randomly chosen subset of the dataset (size specified in parameter). Furthermore, during the training of each tree, the attribute k is randomly chosen when splitting a node. Finally, predictions made by each tree are aggregated to produce a global prediction at the end: the average of the values estimated by all the trees is used in the regression case.

The model defines some parameters, called hyper-parameters, to guide the construction of trees. Among these parameters, we mention :

- The number of trees *n_estimators*: it defines the number of trees trained by the model. Generally, the higher this number is, the more accurate the predictions will be, but this increases processing cost, especially in large datasets.

- Tree depth *max_depth*: it defines the maximum depth of the tree, i.e., the number of levels starting from the root to the last terminal leaf node. A large depth allows for representing maximum information from the dataset, but the model risks overfitting the data.

- The minimum number of observations per node *min_samples_leaf*: it defines the minimum number of observations required for a terminal node. It represents the minimum size of leaves. For a small size, the model can capture the noise in the data.

- The maximum number of features *max_features*: it gives the number of elements of the list K from which the model chooses the pair (k,t_k) (as explained previously). It allows controlling the randomness of the model. The maximum value is the number of features of the dataset (default value). The larger this number is, the less randomness is introduced in the model,

Figure 5.7: Cross validation (K = 5)

but this generates an additional processing cost during the training (choice of the pair (k,t_k)), especially with a very large number of trees.

Generally, these parameters control the growth of the trees. Therefore, opting for deeper trees with small leaves provides a better representation of the data, but the model will probably overfit the data. However, restricting the growth of these trees creates a generalized model, but small trees with large leaves can lead to underfitting (when the model only performs well on training data but performs poorly on testing data) data and thus degrading model performance. Therefore, choosing these parameters is a crucial step in model training for a better-performing generalized model.

One way to calibrate these parameters is to train various models with different values and choose the combination with the best accuracy.
In order to minimize the processing overhead of exploring multiple possibilities, we first use a randomized search using the *RandomizedSearchCV* technique to find an initial value for each parameter (from a large range of input values). Then, we use a grid search using *GridSearchCV* to try all possible combinations (from small ranges of values that bound the previously found values) to find the best combination with the best accuracy. These techniques are based on cross-validation, which uses the Training-Set for training and validation. Indeed, the training subset is randomly divided into k distinct blocks. At each iteration (k times), the model reserves a different block for evaluation and performs the training using the other parts ($k-1$ blocks). This avoids using a part of the dedicated dataset for validation (validation set). Figure 5.7 shows an example of cross-validation with $k = 5$.

5.7 Performance evaluation

The goal is to evaluate the model's capabilities (M_{PRR}) to estimate the packet reception ratio in each square to identify the region's gray zones. We analyze the

results based on the data visualization of the studied scenario. In addition, we identify the strengths and limitations of the proposed model.

We use 75% of the dataset for model training and the remaining 25% for testing, a commonly used ratio. Then, for each input i in the test set X, we compute the corresponding output $\hat{y}_i = M(X_i)$ using the model M (model after the training phase). Then we compare this output with the real value y_i. In this way, we compute the prediction accuracy of each model using the performance metrics presented below.

5.7.1 Performance metrics

To evaluate the accuracy of the proposed ML performance models, we consider two evaluation metrics:

- The prediction **score** R^2 (Eq. 5.4): it represents which part of the variance in the dependent variable is predictable from the independent variables. In other words, it is the proportion of correctly predicted samples. The most precise regression model would be the one with a relatively high R squared, close to 100 when expressed in percentage. We will represent the score in percentage.

- **Normalized Mean Absolute Error (**$NMAE$**)** (Eq. 5.5): it defines the average of the absolute differences between the estimated and observed values of PRR. We want the $NMAE$ to be as small as possible.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{\sum_{i=1}^{n}(y_i - \bar{y})} \text{ , with } \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{5.4}$$

$$NMAE(y, \hat{y}) = \frac{\frac{\sum_{i=1}^{n}|y_i - \bar{y}|}{n}}{\bar{y}} \text{ , with } \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{5.5}$$

5.7.2 Results

For a subset with n_samples of 10000, where 7500 samples are used to train the random forest model, we obtain a score $R^2 = 90.70\%$ and $NMAE = 5.20\%$ on the 2500 remaining sets. Using all the datasets (52007 samples) generated through NETSIM simulations gives scores around 95%, with $NMAE$ of around 5%. These results prove the ability of the ML-performance approach to provide accurate predictions.

Figure 5.8 shows the real observed test points and their corresponding predictions with the random forest model for $K = 10000$ samples. But, of course, when the subset contains all the initial dataset ($K = 52007$ samples), the prediction accuracy is better, with a score almost equal to 0.95, as shown in Figure 5.9.

The performance of the proposed model mostly meets the needs of a proactive network control based on the estimation of the quality of I2V links on the road,

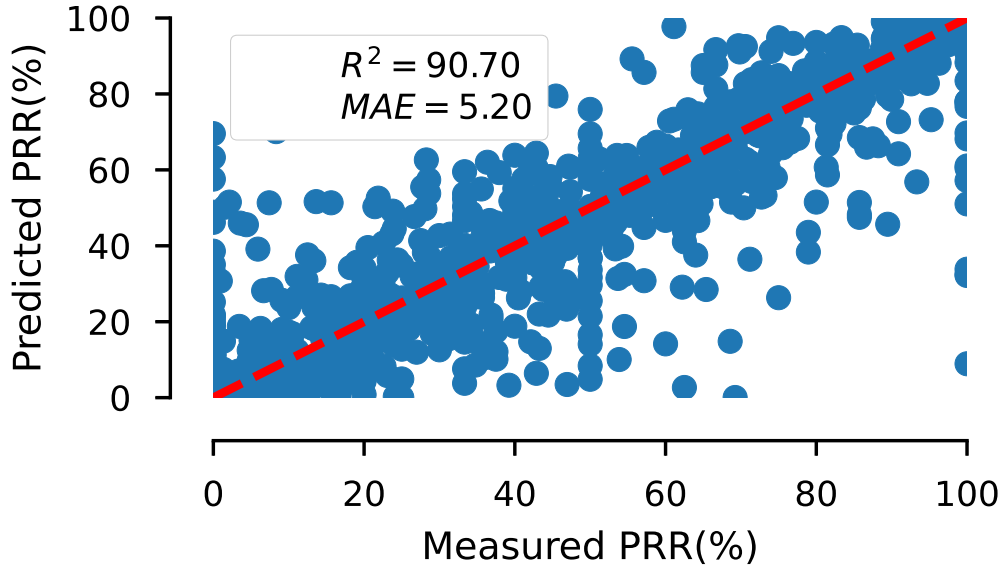


Figure 5.8: Predicted vs. real observations (K = 10000 samples)

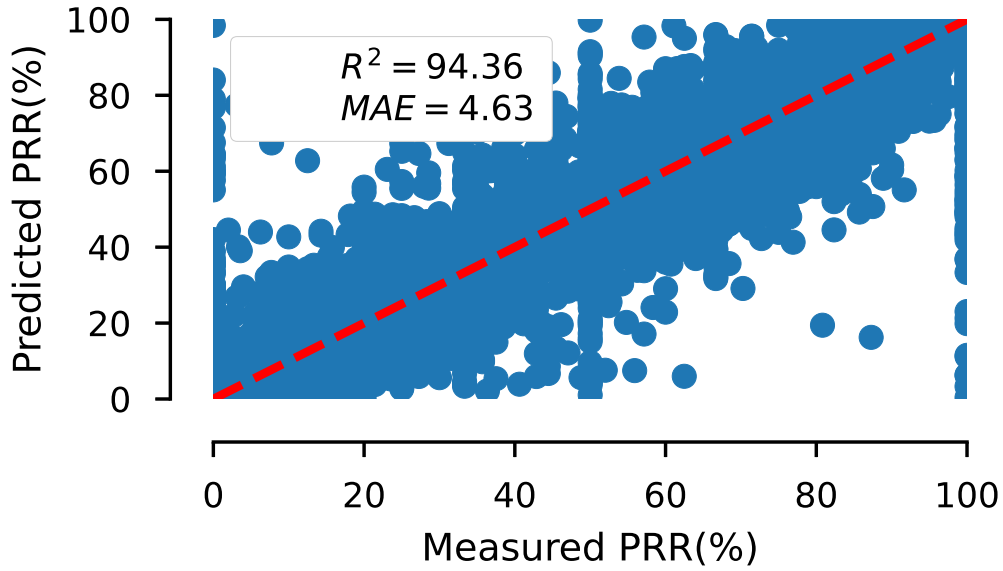


Figure 5.9: Predicted vs. real observations (K = 52007 samples)

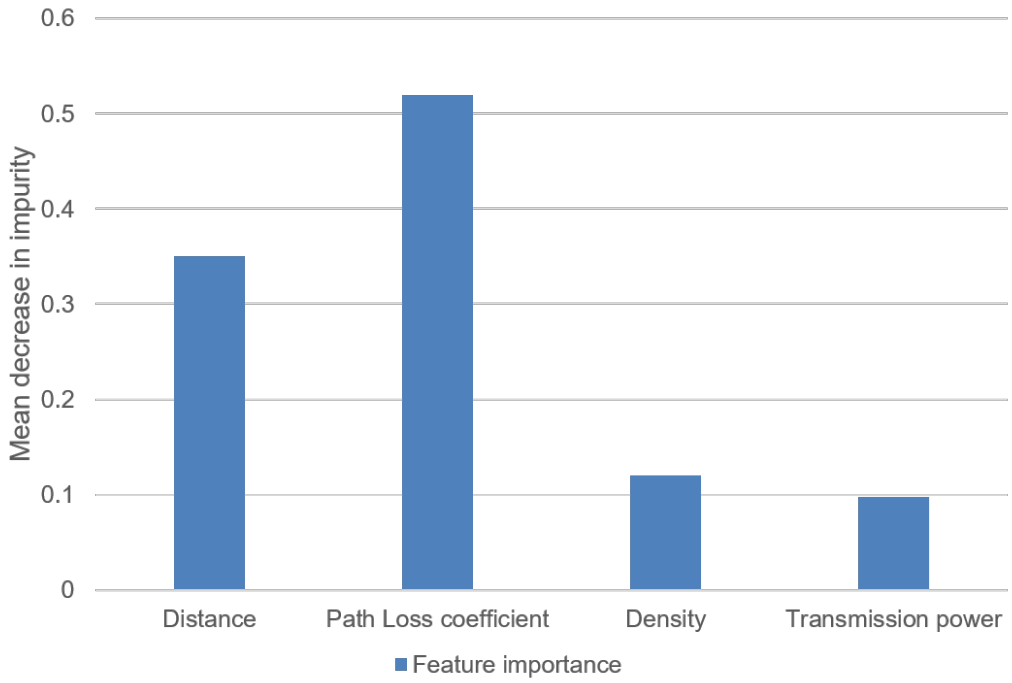


Figure 5.10: Feature importance using Mean Decrease in Impurity (MDI)

which allows defining and updating gray zones periodically in the region in order to update the rebroadcast zones, thus allowing efficient dissemination of alert messages. Now, let us consider the case of a network control function that computes routing paths. Some approaches in the literature consider the quality of a link as a criterion for choosing routing paths. The estimation error (PRR larger or smaller than the estimated duration) may cause an inefficient link selection and, therefore, affect the performance of proactive routing. Most link quality-based routing approaches seek to discard links with low PRR and therefore require a more accurate estimation regarding a poor quality link. The proposed model fully satisfies this point.

Figure 5.10 represents the features' importance, where the Mean Decrease in Impurity (MDI) is used to calculate each feature's importance. It is calculated as the sum over the number of splits (across all trees) that includes the feature, proportionally to the number of samples it splits. This shows that the distance and the packet loss coefficient (where we consider the geographic area, urban/non-urban, the size and presence of the buildings in the area, the weather, etc., that influence communication channels) are the most important feature. Indeed, the absence or disregard of these features will significantly decrease the model score accuracy on the test set (e.g., by removing the path loss from the features, the score accuracy can decrease up to 55%).

(a) Initial RSU placement (b) Q-learning without M_{PRR} (c) Q-learning with M_{PRR}

Figure 5.11: Rebroadcast points placement

5.7.3 Application

The intelligent PRR inference framework (see Figure 5.3) can manage mixed PRR inference with the Q-learning rebroadcast zone placement (described in the previous chapter). Indeed, by separating the network's control plane from its data plane, SDN introduces flexibility in network management and facilitates map updates (gray and rebroadcast zones). Moreover, it simplifies the use of machine learning techniques in the management plane.

In this section, We respectively applied the two proposed techniques, M_{PRR} and Q-learning placement, in order to update the rebroadcast zones periodically according to traffic conditions change (hence the road link's quality and gray zone).

We use the same last scenario (S3) presented in the previous chapter, considering 8 RSUs Figure 5.11a, the same RSUs' position and transmission power, path loss model, and traffic density are initially used to measure the PRR to define the gray zones (where squares with a PRR below 90% are considered gray). First, we estimate the position of the gray zones using M_{PRR} and then apply the Q-learning placement algorithm. As a result, we have the same rebroadcast zone numbers (15 rebroadcast zones as described in the previous chapter results Figure 5.11b) but with a slight difference (1 to 2 squares of shift) in the position of six green circled zones Figure 5.11c. This has practically no impact on the performance of our dissemination procedure, as it remains in the order of a few meters difference and is due to the 4.63% error prediction rate of our model. This shows our PRR model's success in estimating the PRR. Therefore, it is convenient for the SDN controller to update whenever it wants rebroadcast zones' positions considering only some parameters (path loss coefficient, traffic density, and transmission power).

5.8 Synthesis

Estimating link quality on the road paves the way for intelligent and efficient network control. In the proposed approach, we mainly exploited the road packet

reception ratio identification as our model's main learning variable. We chose that this occurs during the update of the rebroadcast zones in our case to ensure efficient and reliable alert messages dissemination, allowing these zones to adapt to the traffic and mobility conditions changes. Indeed, this does not imply any overload on the network. The performance tests showed excellent results in the vast majority of cases.

The model is trained offline using data collected under varying traffic conditions to approximate real mobility and traffic conditions that differ from one hour to another daily. However, the trends captured by the models during real-world training can vary even more when considering new installations and reconstructions in the area (new buildings, facades, parking, etc.), thus changing several parameters such as the communication medium and path loss. Therefore, the controller must be updated by these changes in order to update its road zones and be able to recalculate the path loss by each zone efficiently.

On the other hand, the service provider (ITS) can modify the network parameters in order to optimize its network (modification of the cell coverage, addition, or deletion of a cell). This can affect the model performances. Re-training can occur if the prediction error exceeds a certain threshold (set according to the service using the predictions) with the new data collected and obviously considering these new conditions.

5.9 Conclusion

In this chapter, we presented the prediction model that we propose for road quality link estimation, specifically the PRR on the road. The proposed model is based on supervised machine learning techniques. The main particularity of our approach is that the predictions are triggered while updating the rebroadcast zones in order to adapt to the regular changes in traffic conditions, thus ensuring reliable dissemination of alert/emergency messages.

Due to the lack of datasets meeting the requirements of our study environment and containing the features considered by our model, we generated our dataset. This dataset is derived from a realistic mobility scenario. Training and evaluation of our models are performed using the generated dataset. Performance test results are promising in terms of prediction accuracy.

Even if the experiments are based on DSRC technology, the proposed approach applies to BS entities using cellular networks or any other new network access technology. Indeed, the considered features are independent of any network technology.

This work represents the first step toward developing our scheme for efficient and reliable dissemination of alert messages, which considers all traffic conditions changes and vehicular mobility but not only. Indeed, our model can also be deployed to estimate I2V link quality based on cellular networks in order to meet the requirements of other services, such as multimedia content sharing.

Chapter 6

Conclusion and Perspectives

6.1 Conclusion

The global transportation industry considers C-ITS systems a major technological breakthrough that will revolutionize future mobility. They are counting on deploying these systems to reduce accidents, manage traffic efficiently and improve the driving experience.

Many C-ITS services have been proposed in the literature, combining various interactions between vehicles and their surroundings. Among these services are alert message dissemination services historically based on VANETs. However, these services are very demanding in terms of QoS, and this latter remains a challenge for VANET.

In this context, we have proposed a novel location-based alert message dissemination scheme for Software Defined Vehicular Network (SDVN) to meet low dissemination delay, high-reliability, and full vehicle coverage of such services.

The contributions proposed in this work aim to develop this new dissemination scheme. The summary of the thesis contributions is presented below.

1. For the first contribution, we propose a novel Location-Alert Message Dissemination scheme called LAMD for SDVN that primarily exploits V2I links to widely broadcast alert messages complemented with parsimonious V2V rebroadcasts. V2V rebroadcasts selection is driven by a centralized network controller, which computes rebroadcast locations (or rebroadcast zones). The selection process of vehicles allowed to rebroadcast the message is finally very simple based on their close vicinity to rebroadcast points. We have shown that this alert message dissemination scheme offers highly controlled (ideally full) information coverage while drastically reducing the number of rebroadcasts compared to existing techniques. This considerably reduces the number of collisions and significantly saves network resources. Finally, our proposal induces insignificant network signaling overhead and unimportant vehicle acts.

2. Then, for a second contribution, we improve the performance of our dissemination procedure by optimally placing relay zones. To this end, we propose a new reinforcement learning-based method for optimal dynamic rebroadcast zone placement. This method takes the gray zones of a geographic map under the coverage of the SDN controller as input. It then identifies the number and location of rebroadcast zones. As a result, this method offers an

optimal rebroadcast zone placement by ensuring maximum coverage of the gray zones, which improves the performance of LAMD considerably.

3. In the third contribution, we deploy a machine-learning-based I2V quality link estimation technique on the road to identify gray zones in order to update dynamically rebroadcast zones depending on traffic change conditions. The training and evaluation of this model are done using a dataset that we have generated mainly using the NETSIM framework. The results show a good prediction accuracy rate. This allows updating rebroadcast zones according to regular changes in traffic conditions, thus ensuring reliable dissemination of alert messages.

6.2 Future work

6.2.1 Short-term potentials work

The evaluation results of the proposed dissemination procedure showed significant performance gains and adaptability to traffic fluctuations. However, this work only applies this procedure to alert message dissemination services. One of this project's future developments is to adjust and apply this dissemination strategy to other services, such as multimedia content sharing with different requirements and access technologies (e.g., cellular networks).

Adopting a dynamic approach to place rebroadcast zones according to the gray zones' updates is very promising. However, this technique is currently tested on small maps due to technical constraints. Therefore, we aim to apply our placement and estimation algorithm on large-scale maps by collecting or getting real road traffic data under different traffic conditions.

The link quality estimation technique proposed in the last contribution represents a first step towards developing a global view of the potential state of networks. The proposed approach focuses mainly on I2V links. A short-term perspective of this work is to address the case of V2V links to have a complete view of the potential connectivity of the network nodes.

6.2.2 long-term research directions

As highlighted several times in the thesis, the SDVN architecture positively disrupts the conventional distributed attitude on vehicular networks. It opens a new perspective by enabling a logically centralized network view. Therefore, we are interested in this global view that SDN controllers must build to enable rich and efficient network control. Such a view is established and maintained by the topology discovery service.

The topology discovery service in SDN is a service whose main objective is to build and maintain an up-to-date view of the underlying network topology. This view is primarily used to discover nodes and their links.
A long-term perspective is to combine this service with the quality link estimation

service. This future approach enhances our dissemination procedure to have a relevant view of the network from which the relevant links are selected. Moreover, this view is enriched with the selected links' potential quality. The whole will undoubtedly form a rich representation of the underlying network and will allow for informed and efficient network control, despite the mobility of the vehicles.

Bibliography

[3rd Generation Partnership Project 2015] 3rd Generation Partnership Project. *Study on LTE support for Vehicle to Everything (V2X) services (Release 14)*. Technical Specification Group Services and System Aspects, 2015. 10, 16

[Aldabbagh 2016] Ghadah Aldabbagh, Maaz Rehan, Halabi Hasbullah, Waqas Rehan Omer Chughtai. *A driver safety information broadcast protocol for VANET*. Appl. Math. Inf. Sci, vol. 10, no. 2, pages 451–468, 2016. 31, 33

[Ali 2016] GG Md Nawaz Ali, Peter Han Joo Chong, Syeda Khairunnesa Samantha Edward Chan. *Efficient data dissemination in cooperative multi-RSU vehicular ad hoc networks (VANETs)*. Journal of Systems and Software, vol. 117, pages 508–527, 2016. 17

[Ali 2019] Muhammad Ali, Asad W Malik, Anis U Rahman, Sohail Iqbal Mian M Hamayun. *Position-based emergency message dissemination for Internet of vehicles*. International Journal of Distributed Sensor Networks, vol. 15, no. 7, page 1550147719861585, 2019. 30, 33

[Amoroso 2011] Alessandro Amoroso, Gustavo Marfia Marco Roccetti. *Going realistic and optimal: A distributed multi-hop broadcast algorithm for vehicular safety*. Computer networks, vol. 55, no. 10, pages 2504–2519, 2011. 71

[Araujo 2014] G Araujo, M Queiroz, F Duarte-Figueiredo, A Tostes A Loureiro. *Um protocolo de identificação e minimização de congestionamentos de trafego para redes veiculares*. Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos-(SBRC), 2014. 41

[Arroub 2016] Ayoub Arroub, Bassma Zahi, Essaid Sabir Mohamed Sadik. *A literature review on Smart Cities: Paradigms, opportunities and open problems*. 2016 International conference on wireless networks and mobile communications (WINCOM), pages 180–186. IEEE, 2016. 8

[Aslam 2011] Baber Aslam Cliff C Zou. *Optimal roadside units placement along highways*. 2011 IEEE Consumer Communications and Networking Conference (CCNC), pages 814–815. IEEE, 2011. 52

[Aslam 2012] Baber Aslam, Faisal Amjad Cliff C Zou. *Optimal roadside units placement in urban areas for vehicular networks*. 2012 IEEE symposium on computers and communications (ISCC), pages 000423–000429. IEEE, 2012. 52

[Baccour 2012] Nouha Baccour, Anis Koubâa, Luca Mottola, Marco Antonio Zúñiga, Habib Youssef, Carlo Alberto Boano Mário Alves. *Radio link quality estimation in wireless sensor networks: A survey.* ACM Transactions on Sensor Networks (TOSN), vol. 8, no. 4, pages 1–33, 2012. 70

[Bakhouya 2009] M Bakhouya, J Gaber M Wack. *Performance evaluation of DREAM protocol for inter-vehicle communication.* 2009 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology, pages 289–293. IEEE, 2009. 22

[Balon 2006] Nathan Balon Jinhua Guo. *Increasing broadcast reliability in vehicular ad hoc networks.* Proceedings of the 3rd international workshop on Vehicular ad hoc networks, pages 104–105, 2006. 29, 33

[Bauza 2013] Ramon Bauza, Javier Gozalvez Miguel Sepulcre. *Power-aware link quality estimation for vehicular communication networks.* IEEE Communications Letters, vol. 17, no. 4, pages 649–652, 2013. 71

[Benrhaiem 2020] Wiem Benrhaiem Abdelhakim Senhaji Hafid. *Bayesian networks based reliable broadcast in vehicular networks.* Vehicular Communications, vol. 21, page 100181, 2020. 60, 71

[Bishop 2005] Richard Bishop. *Intelligent Vehicle Technology and Trends Artech house.* Norwood, MA, USA, 2005. 17

[Boban 2017] Mate Boban, Apostolos Kousaridas, Konstantinos Manolakis, Joseph Eichinger Wen Xu. *Use cases, requirements, and design considerations for 5G V2X.* arXiv preprint arXiv:1712.01754, 2017. 42, 61

[Böhm 2010] Annette Böhm, Kristoffer Lidström, Magnus Jonsson Tony Larsson. *Evaluating CALM M5-based vehicle-to-vehicle communication in various road settings through field trials.* IEEE Local Computer Network Conference, pages 613–620. IEEE, 2010. 42, 78

[Breiman 2001] Leo Breiman. *Random forests.* Machine learning, vol. 45, no. 1, pages 5–32, 2001. 75

[Bui 2017] Nicola Bui, Matteo Cesana, S Amir Hosseini, Qi Liao, Ilaria Malanchini Joerg Widmer. *A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques.* IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pages 1790–1821, 2017. 69

[Busanelli 2009] Stefano Busanelli, Gianluigi Ferrari Sooksan Panichpapiboon. *Efficient broadcasting in IEEE 802.11 networks through irresponsible forwarding.* GLOBECOM 2009-2009 IEEE Global Telecommunications Conference, pages 1–6. IEEE, 2009. 22

[Busanelli 2013] Stefano Busanelli, Gianluigi Ferrari, Vito Andrea Giorgio Nicola Iotti. *Information Dissemination in Urban VANETs: Single-Hop or Multi-Hop?* Roadside Networks for Vehicular Communications: Architectures, Applications, and Test Fields, pages 237–263. IGI Global, 2013. 41, 77

[Cai 2014] Xuelian Cai, Ying He, Chunchun Zhao, Lina Zhu Changle Li. *LSGO: Link state aware geographic opportunistic routing protocol for VANETs.* EURASIP Journal on wireless communications and networking, vol. 2014, no. 1, pages 1–10, 2014. 71

[Casademont 2019] Jordi Casademont, Elena Lopez-Aguilera Josep Paradells. *Wake-Up Radio Systems for Cooperative-Intelligent Transport Systems Architecture.* 2019 7th International Conference on Future Internet of Things and Cloud (FiCloud), pages 358–363. IEEE, 2019. ix, 8, 9

[Castiaux 2021] Julien Castiaux. *So you want to solve Freecell? An academic journey into crafting a solitaire game solver.* 2021. 56

[Cavalcante 2012] Evellyn S Cavalcante, André LL Aquino, Gisele L Pappa Antonio AF Loureiro. *Roadside unit deployment for information dissemination in a VANET: An evolutionary approach.* Proceedings of the 14th annual conference companion on Genetic and evolutionary computation, pages 27–34, 2012. 52

[Cerar 2021] Gregor Cerar, Halil Yetgin, Mihael Mohorčič Carolina Fortuna. *Machine learning for wireless link quality estimation: A survey.* IEEE Communications Surveys & Tutorials, vol. 23, no. 2, pages 696–728, 2021. 70, 71

[Cheng 2019] Xiaolu Cheng Baohua Huang. *A center-based secure and stable clustering algorithm for VANETs on highways.* Wireless communications and mobile computing, vol. 2019, 2019. 30, 33

[Ciccarese 2009] Giovanni Ciccarese, Mario De Blasi, Pierluigi Marra, Cosimo Palazzo Luigi Patrono. *On the use of control packets for intelligent flooding in VANETs.* 2009 IEEE Wireless Communications and Networking Conference, pages 1–6. IEEE, 2009. 40

[Committee 1999] IEEE Computer Society LAN MAN Standards Committee*et al. Wireless LAN medium access control (MAC) and physical layer (PHY) specifications.* ANSI/IEEE Std. 802.11-1999, 1999. 23

[de Sousa 2015] R de Sousa A Soares. *Estimativa e sinalizaç ao de congestionamentos de tráfego através de redes veiculares v2v.* Simpósio Brasileiro de Redes de Computadores, SBRC, page 1, 2015. 41

[de Sousa 2018] Roniel S de Sousa, Felipe S da Costa, Andre CB Soares, Luiz FM Vieira Antonio AF Loureiro. *Geo-sdvn: A geocast protocol for software*

defined vehicular networks. 2018 IEEE International Conference on Communications (ICC), pages 1–6. IEEE, 2018. 41

[Dressler 2011] Falko Dressler, Frank Kargl, Jorg Ott, Ozan K Tonguz Lars Wischhof. *Research challenges in intervehicular communication: lessons of the 2010 Dagstuhl Seminar.* IEEE Communications Magazine, vol. 49, no. 5, pages 158–164, 2011. 17, 18

[ETSI] TR ETSI. *102 962 V1. 1.1 (Feb. 2012)"Intelligent Transport Systems (ITS); Framework for Public Mobile Networks in Cooperative ITS (C-ITS)".* , Technical Report, European Telecommunications Standards Institute (ETSI 18

[ETSI 2009] TCITS ETSI. *ETSI TS 102 638 v1. 1.1-intelligent transport systems (ITS); vehicular communications; basic set of applications; definitions.* Standard, TC ITS, 2009. 10, 11, 12, 13, 14, 20, 24

[ETSI 2013] TCITS ETSI. *Intelligent Transport Systems; Vehicular Communications; Basic Set of Applications; Part 3: Specification of Decentralized Environmental Notification Basic Service, Std.* ETSI EN Std, vol. 302, pages 637–3, 2013. 13, 24

[ETSI 2014] TCITS ETSI. *Intelligent Transport Systems; Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service, Std.* ETSI EN Std, vol. 302, pages 637–2, 2014. 13

[ETSI 2019] TCITS ETSI. *Intelligent Transport Systems; ITS-G5 Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band.* ETSI EN Std, vol. 302, 2019. 15, 24

[ETSI 2020] ETSI. *ITS definition.* `https://www.etsi.org/images/files/ETSITechnologyLeaflets/IntelligentTra%20nsportSystems.pdf`, 2020. 8, 10

[Feamster 2014] Nick Feamster, Jennifer Rexford Ellen Zegura. *The road to SDN: an intellectual history of programmable networks.* ACM SIGCOMM Computer Communication Review, vol. 44, no. 2, pages 87–98, 2014. 75

[Festag 2014] Andreas Festag. *Cooperative intelligent transport systems standards in Europe.* IEEE communications magazine, vol. 52, no. 12, pages 166–172, 2014. xi, 14

[Försterling 2015] Frank Försterling. *Electronic Horizon How the Cloud improves the connected vehicle*, 2015. 1

[Foster 2020] Nate Foster, Nick McKeown, Jennifer Rexford, Guru Parulkar, Larry Peterson Oguz Sunay. *Using deep programmability to put network owners in control*, 2020. 75

[Foundation 2012] Open Networking Foundation. *Software-defined networking: The new norm for networks.* ONF White Paper, vol. 2, pages 2–6, 2012. ix, 20

[Garip 2015] Mevlut Turker Garip, Mehmet Emre Gursoy, Peter Reiher Mario Gerla. *Scalable reactive vehicle-to-vehicle congestion avoidance mechanism.* 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC), pages 943–948. IEEE, 2015. 41

[Gawas 2017] Mahadev A Gawas, Palash Hurkat, Varun Goyal Lucy J Gudino. *Cross layer approach for efficient dissemination of emergency messages in VANETs.* 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), pages 206–211. IEEE, 2017. 30, 33

[Ge 2017] Xiaohu Ge, Zipeng Li Shikuan Li. *5G software defined vehicular networks.* IEEE Communications Magazine, vol. 55, no. 7, pages 87–93, 2017. 17

[Gerharz 2003] Michael Gerharz, Christian de Waal, Matthias Frank Paul James. *A practical view on quality-of-service support in wireless ad hoc networks.* Proc. of the 3rd IEEE Workshop on Applications and Services in Wireless Networks (ASWN), pages 185–196. Citeseer, 2003. 23

[Gozálvez 2012] Javier Gozálvez, Miguel Sepulcre Ramon Bauza. *IEEE 802.11 p vehicle to infrastructure communications in urban environments.* IEEE Communications Magazine, vol. 50, no. 5, pages 176–183, 2012. 42, 78

[Hameed Mir 2014] Zeeshan Hameed Mir Fethi Filali. *LTE and IEEE 802.11 p for vehicular networking: a performance evaluation.* EURASIP Journal on Wireless Communications and Networking, vol. 2014, no. 1, pages 1–15, 2014. 18, 19

[Hassan 2011] Md Imrul Hassan, Hai L Vu Taka Sakurai. *Performance analysis of the IEEE 802.11 MAC protocol for DSRC safety applications.* IEEE Transactions on vehicular technology, vol. 60, no. 8, pages 3882–3896, 2011. 15, 18

[He 2016] Zongjian He, Jiannong Cao Xuefeng Liu. *SDVN: Enabling rapid network innovation for heterogeneous vehicular communication.* IEEE network, vol. 30, no. 4, pages 10–15, 2016. 2, 19

[Ibrahim 2020] Banar Fareed Ibrahim, Mehmet Toycan Hiwa Abdulkarim Mawlood. *A comprehensive survey on VANET broadcast protocols.* 2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM), pages 298–302. IEEE, 2020. 28

[Jerbi 2011] Moez Jerbi, Sidi-Mohammed Senouci, Yacine Ghamri-Doudane Mohamed Cherif. *Vehicular communications networks: Current trends and*

challenges. Next Generation Mobile Networks and Ubiquitous Computing, pages 251–262. IGI Global, 2011. 18

[Katsaros 2013] Konstantinos Katsaros, Mehrdad Dianati Long Le. *Effective implementation of location services for VANETs in hybrid network infrastructures*. 2013 IEEE International Conference on Communications Workshops (ICC), pages 521–525, 2013. 25

[Kawser 2019] Mohammad T Kawser, Md Saymon Fahad, Sakib Ahmed, S Safwan Sajjad Hasib A Rafi. *The perspective of vehicle-to-everything (v2x) communication towards 5g*. IJCSNS, vol. 19, no. 4, page 146, 2019. ix, 10

[Kenney 2011] John B Kenney. *Dedicated short-range communications (DSRC) standards in the United States*. Proceedings of the IEEE, vol. 99, no. 7, pages 1162–1182, 2011. 24

[KL 2017] Kushan Sudheera KL, Maode Ma Peter Han Joo Chong. *Efficient flow instantiation via source routing in software defined vehicular networks*. 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), pages 1–5. IEEE, 2017. 7

[Klaine 2017] Paulo Valente Klaine, Muhammad Ali Imran, Oluwakayode Onireti Richard Demo Souza. *A survey of machine learning techniques applied to self-organizing cellular networks*. IEEE Communications Surveys & Tutorials, vol. 19, no. 4, pages 2392–2431, 2017. 74

[Koenig 1992] Sven Koenig Reid G Simmons. *Complexity analysis of real-time reinforcement learning applied to finding shortest paths in deterministic domains*. , CARNEGIE-MELLON UNIV PITTSBURGH PA SCHOOL OF COMPUTER SCIENCE, 1992. 60

[Korkmaz 2004] Gökhan Korkmaz, Eylem Ekici, Füsun Özgüner Ümit Özgüner. *Urban multi-hop broadcast protocol for inter-vehicle communication systems*. Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks, pages 76–85, 2004. 21

[Kotsiantis 2007] Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas*et al.* *Supervised machine learning: A review of classification techniques*. Emerging artificial intelligence applications in computer engineering, vol. 160, no. 1, pages 3–24, 2007. 73, 74

[Ku 2014] Ian Ku, You Lu, Mario Gerla, Rafael L Gomes, Francesco Ongaro Eduardo Cerqueira. *Towards software-defined VANET: Architecture and services*. 2014 13th annual Mediterranean ad hoc networking workshop (MED-HOC-NET), pages 103–110. IEEE, 2014. 17, 19

[Kuffermann 2014] Kyle Kuffermann. An implementation of the ieee1609. 4 wave standard for use in a vehicular networking testbed. Florida Atlantic University, 2014. 35

[Lee 2010] Uichin Lee Mario Gerla. *A survey of urban vehicular sensing platforms.* Computer Networks, vol. 54, no. 4, pages 527–544, 2010. 22

[Lindeberg 1998] Tony Lindeberg. *Feature detection with automatic scale selection.* International journal of computer vision, vol. 30, no. 2, pages 79–116, 1998. 30

[Liu 2007] Pei Liu, Zhifeng Tao, Sathya Narayanan, Thanasis Korakis Shivendra S Panwar. *CoopMAC: A cooperative MAC for wireless LANs.* IEEE journal on selected areas in Communications, vol. 25, no. 2, pages 340–354, 2007. 22

[Liu 2013] Yazhi Liu, Jianwei Niu, Jian Ma Wendong Wang. *File downloading oriented roadside units deployment for vehicular networks.* Journal of Systems Architecture, vol. 59, no. 10, pages 938–946, 2013. 52

[Liu 2015] Bingyi Liu, Dongyao Jia, Jianping Wang, Kejie Lu Libing Wu. *Cloud-assisted safety message dissemination in VANET–cellular heterogeneous wireless network.* ieee systems journal, vol. 11, no. 1, pages 128–139, 2015. 29, 33

[Liu 2018a] Lei Liu, Chen Chen, Tie Qiu, Mengyuan Zhang, Siyu Li Bin Zhou. *A data dissemination scheme based on clustering and probabilistic broadcasting in VANETs.* Vehicular Communications, vol. 13, pages 78–88, 2018. 30, 33

[Liu 2018b] Sheng Liu, Weidong Xiang M Xavier Punithan. *An empirical study on performance of DSRC and LTE-4G for vehicular communications.* 2018 IEEE 88th Vehicular Technology Conference (VTC-Fall), pages 1–5. IEEE, 2018. 18, 19

[Lowrance 2017] Christopher J Lowrance Adrian P Lauf. *Link quality estimation in ad hoc and mesh networks: A survey and future directions.* Wireless Personal Communications, vol. 96, no. 1, pages 475–508, 2017. 71

[Ma 2009] Xiaomin Ma, Xianbo Chen Hazem H Refai. *Performance and reliability of DSRC vehicular safety communication: a formal analysis.* EURASIP Journal on Wireless Communications and Networking, vol. 2009, pages 1–13, 2009. 15, 18

[Mao 2018] Qian Mao, Fei Hu Qi Hao. *Deep learning for intelligent wireless networks: A comprehensive survey.* IEEE Communications Surveys & Tutorials, vol. 20, no. 4, pages 2595–2621, 2018. 71

[Maowad 2012] Hafez Maowad Eman Shaaban. *Enhancing AOMDV routing protocol for V2V communication.* Proceedings of the 6th International Conference on Communications and Information Technology, and Proceedings of the 3rd World Conference on Education and Educational Technologies, pages 20–27, 2012. 31, 33

[Mehar 2015] Sara Mehar, Sidi Mohammed Senouci, Ali Kies Mekkakia Maaza Zoulikha. *An optimized roadside units (RSU) placement for delay-sensitive applications in vehicular networks.* 2015 12th Annual IEEE consumer communications and networking conference (CCNC), pages 121–127. IEEE, 2015. 52

[Meireles 2010] Rui Meireles, Mate Boban, Peter Steenkiste, Ozan Tonguz Joao Barros. *Experimental study on the impact of vehicular obstructions in VANETs.* 2010 IEEE Vehicular Networking Conference, pages 338–345. IEEE, 2010. 42, 72

[Meneguette 2014] Rodolfo Ipolito Meneguette, Azzedine Boukerche, Guilherme Maia, Antonio AF Loureiro Leandro A Villas. *A self-adaptive data dissemination solution for intelligent transportation systems.* Proceedings of the 11th ACM symposium on Performance evaluation of wireless ad hoc, sensor, & ubiquitous networks, pages 69–76, 2014. 42

[Milojevic 2014] Milos Milojevic Veselin Rakocevic. *Distributed road traffic congestion quantification using cooperative VANETs.* 2014 13th annual Mediterranean ad hoc networking workshop (MED-HOC-NET), pages 203–210. IEEE, 2014. 41

[Narang 2015] Ghanishtha Narang Yogesh Juneja. *Review on classification of different VANET Protocols based on routing information.* International Journal of Advanced Research in Computer and Communication Engineering, vol. 4, no. 5, pages 388–392, 2015. 31

[Naumov 2006] Valery Naumov, Rainer Baumann Thomas Gross. *An evaluation of inter-vehicle ad hoc networks based on realistic vehicular traces.* Proceedings of the 7th ACM international symposium on Mobile ad hoc networking and computing, pages 108–119, 2006. 31, 33

[Network 2015] Evolved Universal Terrestrial Radio Access Network. *ATIS 3GPP SPECIFICATION.* 2015. 10

[Ning 2019] Zhaolong Ning, Jun Huang Xiaojie Wang. *Vehicular fog computing: Enabling real-time traffic management for smart cities.* IEEE Wireless Communications, vol. 26, no. 1, pages 87–93, 2019. 25

[Okada 2009] Hiraku Okada, Akira Takano Kenichi Mase. *A proposal of link metric for next-hop forwarding methods in vehicular ad hoc networks.* 2009

6th IEEE Consumer Communications and Networking Conference, pages 1–5. IEEE, 2009. 71

[Oliveira 2017] Renê Oliveira, Carlos Montez, Azzedine Boukerche Michelle S Wangham. *Reliable data dissemination protocol for VANET traffic safety applications.* Ad Hoc Networks, vol. 63, pages 30–44, 2017. 28, 33, 40, 41

[ONISR 2022] ONISR. `https://www.onisr.securite-routiere.gouv.fr/`, Accessed : 18-08-2022. 1

[Palazzi 2012] Claudio E Palazzi, Fabio Pezzoni Pedro M Ruiz. *Delay-bounded data gathering in urban vehicular sensor networks.* Pervasive and Mobile Computing, vol. 8, no. 2, pages 180–193, 2012. 21

[Palma 2012] Adriano Palma, Paulo Rogério Pereira, Paulo Rogério Pereira Augusto Casaca. *Multicast routing protocol for vehicular delay-tolerant networks.* 2012 IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pages 753–760. IEEE, 2012. 22

[Panichpapiboon 2011] Sooksan Panichpapiboon Wasan Pattara-Atikom. *A review of information dissemination protocols for vehicular ad hoc networks.* IEEE Communications Surveys & Tutorials, vol. 14, no. 3, pages 784–798, 2011. 44, 65

[Peng 2007] Jun Peng Liang Cheng. *A distributed MAC scheme for emergency message dissemination in vehicular ad hoc networks.* IEEE Transactions on Vehicular Technology, vol. 56, no. 6, pages 3300–3308, 2007. 22

[Puterman 2014] Martin L Puterman. Markov decision processes: discrete stochastic dynamic programming. John Wiley & Sons, 2014. 55

[Rahim 2009] Aneel Rahim, Mehboob Yasin, Imran Ahmad, Zeeshan Shafi Khan Muhammad Sher. *Relevance based approach with virtual queue for vehicular adhoc networks.* 2009 2nd International Conference on Computer, Control and Communication, pages 1–4. IEEE, 2009. 23

[Rak 2013] Jacek Rak. *LLA: A new anypath routing scheme providing long path lifetime in VANETs.* IEEE communications letters, vol. 18, no. 2, pages 281–284, 2013. 71

[Ramya 2019] Panthangi M Ramya, Mate Boban, Chan Zhou Sławomir Stańczak. *Using learning methods for v2v path loss prediction.* 2019 IEEE Wireless Communications and Networking Conference (WCNC), pages 1–6. IEEE, 2019. 71

[Rasool 2017] Illa Ul Rasool, Yousaf Bin Zikria Sung Won Kim. *A review of wireless access vehicular environment multichannel operational medium*

access control protocols: Quality-of-service analysis and other related issues. International Journal of Distributed Sensor Networks, vol. 13, no. 5, page 1550147717710174, 2017. 15

[Rezaei 2015] Ali Asghar Rezaei. *Tiling fullerene surface with heptagon and octagon.* Fullerenes, Nanotubes and Carbon Nanostructures, vol. 23, no. 12, pages 1033–1036, 2015. 43

[Sagi 2018] O Sagi L Rokach. *Ensemble learning: A survey Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 2018. 74

[Salahuddin 2014] Mohammad Ali Salahuddin, Ala Al-Fuqaha Mohsen Guizani. *Software-defined networking for rsu clouds in support of the internet of vehicles.* IEEE Internet of Things journal, vol. 2, no. 2, pages 133–144, 2014. 2

[Salim 2007] Flora Dilys Salim, Seng Wai Loke, Andry Rakotonirainy, Bala Srinivasan Shonali Krishnaswamy. *Collision pattern modeling and real-time collision detection at road intersections.* 2007 ieee intelligent transportation systems conference, pages 161–166. IEEE, 2007. 2, 24, 27

[Sanguesa 2016] Julio A Sanguesa, Manuel Fogue, Piedad Garrido, Francisco J Martinez, Juan-Carlos Cano Carlos T Calafate. *A survey and comparative study of broadcast warning message dissemination schemes for VANETs.* Mobile Information Systems, vol. 2016, 2016. 28

[Sato 2020] Goshi Sato, Akira Sakuraba, Noriki Uchida Yoshitaka Shibata. *A new road state information platform based on crowed sensing on challenged network environments.* Internet of Things, page 100214, 2020. 10

[Shahwani 2021] Hamayoun Shahwani, Syed Attique Shah, Muhammad Ashraf, Muhammad Akram, Jaehoon Paul Jeong Jitae Shin. *A comprehensive survey on data dissemination in Vehicular Ad Hoc Networks.* Vehicular Communications, page 100420, 2021. 28

[Slavik 2010] M Slavik I Mahgoub. *Stochastic broadcast for vanet, CCNC'10: Proceedings of the 7th IEEE conference on Consumer communications and networking conference*, 2010. 21

[Sudheer 2017] KL Kushan Sudheer, Maode Ma Peter Han Joo Chong. *Link dynamics based packet routing framework for software defined vehicular networks.* GLOBECOM 2017-2017 IEEE Global Communications Conference, pages 1–6. IEEE, 2017. 7

[Sun 2000] Min-Te Sun, Wu-Chi Feng, Ten-Hwang Lai, Kentaro Yamada, Hiromi Okada Kikuo Fujimura. *GPS-based message broadcasting for inter-vehicle communication.* Proceedings 2000 International Conference on Parallel Processing, pages 279–286. IEEE, 2000. 21

[Sun 2010] Yipin Sun, Xiaodong Lin, Rongxing Lu, Xuemin Shen Jinshu Su. *Roadside units deployment for efficient short-time certificate updating in VANETs.* 2010 IEEE international conference on communications, pages 1–5. IEEE, 2010. 52

[Sutton 2018] Richard S Sutton Andrew G Barto. Reinforcement learning: An introduction. MIT press, 2018. 58, 59

[Tang 2007] Lei Tang, Kuang-Ching Wang, Yong Huang Fangming Gu. *Channel characterization and link quality assessment of IEEE 802.15. 4-compliant radio for factory environments.* IEEE Transactions on industrial informatics, vol. 3, no. 2, pages 99–110, 2007. 70

[Tian 2016] Bin Tian. *Data dissemination protocols and mobility model for VANETs.* PhD thesis, Université Blaise Pascal-Clermont-Ferrand II, 2016. 42

[TOMTOM 2022] TOMTOM. `https://www.tomtom.com/fr_fr/`, Accessed : 18-08-2022. 1

[Torrent-Moreno 2004] Marc Torrent-Moreno, Daniel Jiang Hannes Hartenstein. *Broadcast reception rates and effects of priority access in 802.11-based vehicular ad-hoc networks.* Proceedings of the 1st ACM international workshop on Vehicular ad hoc networks, pages 10–18, 2004. 23

[Toufga 2017] Soufian Toufga, Philippe Owezarski, Slim Abdellatif Thierry Villemur. *An SDN hybrid architecture for vehicular networks: Application to Intelligent Transport System.* arXiv preprint arXiv:1712.05307, 2017. 2, 25

[Trullols 2009] O Trullols, José M Barcelo-Ordinas, Marco Fiore, Claudio Casetti C-F Chiasserini. *A max coverage formulation for information dissemination in vehicular networks.* 2009 IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, pages 154–160. IEEE, 2009. 52

[Tseng 2002] Yu-Chee Tseng, Sze-Yao Ni, Yuh-Shyan Chen Jang-Ping Sheu. *The broadcast storm problem in a mobile ad hoc network.* Wireless networks, vol. 8, no. 2, pages 153–167, 2002. 2, 21, 27

[Turan 2021] Bugra Turan, Ali Uyrus, O Nuri Koc, Emrah Kar Sinem Coleri. *Machine Learning Aided Path Loss Estimator and Jammer Detector for Heterogeneous Vehicular Networks.* 2021 IEEE Global Communications Conference (GLOBECOM), pages 1–6. IEEE, 2021. 71

[Ullah 2018] Ata Ullah, Shumayla Yaqoob, Muhammad Imran Huansheng Ning. *Emergency message dissemination schemes based on congestion avoidance*

in VANET and vehicular FoG computing. IEEE Access, vol. 7, pages 1570–1585, 2018. 28

[Virdaus 2017] Irvanda Kurniadi Virdaus, Moonsoo Kang, Seokjoo Shin, Chung Ghiu Lee Jae-Young Pyun. *A counting-based broadcast model of emergency message dissemination in VANETs.* 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), pages 927–930. IEEE, 2017. 29, 33

[Wang 2012] Hao Wang, Guo-zhen Tan Ji-xiang Yang. *An improved VANET intelligent forward decision-making routing algorithm.* Journal of Networks, vol. 7, no. 10, page 1546, 2012. 71

[Wang 2015] Chu-Fu Wang, Yan-Ping Chiou Guan-Hsiung Liaw. *Nexthop selection mechanism for nodes with heterogeneous transmission range in VANETs.* Computer Communications, vol. 55, pages 22–31, 2015. 71

[Wang 2017] Mowei Wang, Yong Cui, Xin Wang, Shihan Xiao Junchen Jiang. *Machine learning for networking: Workflow, advances and opportunities.* Ieee Network, vol. 32, no. 2, pages 92–99, 2017. ix, 74, 75, 77

[Watkins 1992] Christopher JCH Watkins Peter Dayan. *Q-learning.* Machine learning, vol. 8, no. 3, pages 279–292, 1992. 58

[Weber 2021] Julia Silva Weber, Miguel Neves Tiago Ferreto. *VANET simulators: an updated review.* Journal of the Brazilian Computer Society, vol. 27, no. 1, pages 1–31, 2021. 40, 78

[Wegner 2018] Martin Wegner, Timo Schwarz Lars Wolf. *Connectivity Maps for V2I communication via ETSI ITS-G5.* 2018 IEEE Vehicular Networking Conference (VNC), pages 1–8. IEEE, 2018. 12

[Wisitpongphan 2007] Nawaporn Wisitpongphan, Ozan K Tonguz, Jayendra S Parikh, Priyantha Mudalige, Fan Bai Varsha Sadekar. *Broadcast storm mitigation techniques in vehicular ad hoc networks.* IEEE Wireless Communications, vol. 14, no. 6, pages 84–94, 2007. 21

[Wu 2017] Libing Wu, Lei Nie, Jing Fan, Yanxiang He, Qin Liu Dan Wu. *An efficient multi-hop broadcast protocol for emergency messages dissemination in VANETs.* Chinese Journal of Electronics, vol. 26, no. 3, pages 614–623, 2017. 29, 33

[Xu 2017] Zhigang Xu, Xiaochi Li, Xiangmo Zhao, Michael H Zhang Zhongren Wang. *DSRC versus 4G-LTE for connected vehicle applications: A study on field experiments of vehicular communication performance.* Journal of advanced transportation, vol. 2017, 2017. 7, 36

[Ye 2018] Hao Ye, Le Liang, Geoffrey Ye Li, JoonBeom Kim, Lu Lu May Wu. *Machine learning for vehicular networks: Recent advances and application examples.* ieee vehicular technology magazine, vol. 13, no. 2, pages 94–101, 2018. 71

[Zakaria 2021] Yahia A Zakaria, Ehab KI Hamad, AS Abd Elhamid Kamel M El-Khatib. *Propagation measurements and calculation of path loss exponent for outdoor cellular communication systems at 3.5 GHz.* Radioelectronics and Communications Systems, vol. 64, no. 5, pages 247–254, 2021. 78

[Zhang 2018] Ziyao Zhang, Liang Ma, Kin K Leung, Leandros Tassiulas Jeremy Tucker. *Q-placement: Reinforcement-learning-based service placement in software-defined networks.* 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS), pages 1527–1532. IEEE, 2018. 61