Sidelink Radio Resource Management for Vehicular Communications in Co-Operative Automated Driving

Dissertation zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)

Vorgelegt im Fachbereich Elektrotechnik/Informatik
der Universität Kassel

von Prajwal Makkimane Keshavamurthy

Erstgutachter: Prof. Dr. sc. techn. Dirk Dahlhaus
Zweitgutachter: Prof. Dr.-Ing. habil. Falko Dressler

Disputation am 26. August 2024

Acknowledgement

At first, I express my deep gratitude to Prof. Dr. sc. techn. Dirk Dahlhaus for mentoring me. This thesis would not have been possible without his valuable advices, unwavering support, timely feedback and immense patience.

My sincere thanks to Huawei German Research Center, Munich, for the incredible doctoral research opportunity. My special thanks to Dr. Emmanouil Pateromichelakis, Dr.-Ing. Chan Zhou and Dr. Panagiotis Spapis for their open and enthusiastic collaborations. I extend my special thanks to Prof. Dr. Alexandros Kaloxylos for his encouragement.

Finally, I acknowledge with gratitude the incredible support of my wife Kavya Danivas. Also, I am eternally grateful to my parents Rajalakshmi K. P. and Keshavamurthy M. K. and my sister Shubha M. K. for their love and support.

Kurzfassung

Automatisiertes Fahren zeichnet sich als zentraler Technologietrend ab und bietet eine Chance zur Koordinierung autonomer Fahrzeuge zur Verbesserung des Straßenverkehrsflusses, der Kraftstoffeffizienz und der Betriebssicherheit. In diesem Zusammenhang stellt kooperatives automatisiertes Fahren (KAF) ein Mittel für Fahrzeuge dar, eine kooperative Planung der Bewegungspfade durchzuführen und die Umgebungswahrnehmung durch Nutzung ihrer Fahrzeug-zu-Fahrzeug (FZF)-Kommunikationsfunktionen mittels in Kommunikationsdistanz befindlichen Fahrzeugbenutzerequipments (FBEs) zu verbessern. Typische Beispiele für KAF sind das Kolonnenfahren, ein kooperativer Spurwechsel und die kooperative Kollisionsvermeidung. Die Anforderungen an die Dienstgüte (DG) für die KAF-FZF-Kommunikation sind sehr unterschiedlich mit einer Zuverlässigkeit von 90 bis 99,999 Prozent und einer Ende-zu-Ende-Paketverzögerung von 3 bis 100 ms. Hierbei spiegelt die Flexibilität der DG-Anforderungen die Leistungstoleranz wider, die sich je nach Anwendungsfall und unterschiedlichen Niveaus der Automatisierung auf Kosten suboptimaler Abläufe einstellen.

Eine Schlüsselrolle spielt das Radio Resource Management (RRM) im sogenannten Sidelink, also in der direkten FZF-Übertragung, mit DG-bewusster Funkressourcenplanung (FRP) zur Erfüllung der KAF-DG-Anforderungen. Letztere bedingen eine enge Interaktion zwischen KAF-Anwendungen und dem Sidelink-RRM. Darüber hinaus erfordert KAF FZF-Interaktionen und Interferenzkoordination zwischen FBEs unterschiedlicher Mobilfunknetzbetreiber (MNB), und vertikale Partner (z. B. Automobilunternehmen) möchten möglicherweise die Sidelink-Funkressourcen steuern. Daher muss das Sidelink-RRM vom Netzwerk des Be-

treibers abstrahiert oder entkoppelt werden. Zu diesem Zweck wird die Sidelink-RRM-Funktionalität auf betreiberunabhängige cloudbasierte Netzwerkeinheiten verlagert, auf die FBEs über unterschiedliche MNBs hinweg zugreifen können.

Es wird ein cloud-fähiges Sidelink-RRM-Framework eingeführt, das die Platzierung der Sidelink-RRM-Funktionalität in einer logisch zentralisierten Cloud-Einheit und in einer Umgebung mit unterschiedlichen Betreibern ermöglicht. Durch die Berücksichtigung der unterschiedlichen DG-Anforderungen einer KAF-Multicast-Gruppenkommunikation in Form von Nutzenfunktionen wird ein DG-abhängiges und nutzenbasiertes mehrkriterielles Optimierungsproblem zur Funkressourcenzuteilung beschrieben. Letzteres zielt auf die Maximierung einer aggregierten Nutzenfunktion ab, wobei sich die Ressourcen auf Ressourcenblöcke (RBs) und Modulations- und Kodierungsschemata beziehen. Darüber hinaus wird eine heuristische zentralisierte FRP-Lösung geringer Komplexität vorgeschlagen und für cloudbasiertes Sidelink-RRM mit potenziell hoher Fahrzeug-zu-Netzwerk (FZN)-Verbindungsverzögerung sowie für dynamische Fahrzeugumgebungen analysiert. Um das Problem der Erfassung fehlender Kanalzustandsinformation (KZI) zu adressieren, wird die dynamische FRP-Lösung unter Verwendung langsam variierender Kanalparameter (z. B. der Pfaddämpfung), die als langsame KZI bezeichnet werden, verbessert.

Ein Edge-Cloud-fähiges semi-zentralisiertes Sidelink-RRM-Framework wird vorgeschlagen, um den dynamischen Scheduler unter Ausnutzung der Rechenleistung der autonomen Fahrzeuge näher an die FBEs zu bringen. Um die resultierenden Vorteile zu nutzen, wird das Ressourcenallokationsproblem in drei Aufgaben mit hoher Modularität und Abstimmbarkeit aufgeteilt. Diese Aufgaben umfassen eine FBE-Clusterbildung, eine RB-Pool-Zuweisung zwischen Clustern und eine Ressourcenzuweisung zur optimierten Sidelink-FRP innerhalb eines Clusters, welche auf unterschiedlichen Zeitskalen an verschiedenen Cloud-Entitäten (zentral und am

Rand) mit einer reduzierten dynamischen Planungsverzögerung durchgeführt werden können. Ein graphentheoretischer Ansatz wird vorgestellt mit

1. einem KAF-Fahrzeuggruppen-Clustering als Clique-Partitionierungsproblem, das die Koordination auf der Steuerungsebene zwischen sendenden Fahrzeugen ermöglicht, um das Hidden-Node-Problem und die durch die Halbduplex-Einschränkung auferlegten Einschränkungen zu vermeiden und

2. einer Cluster-zu-RB-Pool-Zuordnung als Max-Min-Fairness-Problem auf einem gewichteten Ressourcenkonfliktgraphen, das den Kompromiss zwischen einer verbesserten Spektrumsnutzung (durch die effiziente Wiederverwendung von Sidelink-Ressourcen zwischen Clustern) und der Begrenzung der Interferenzen zwischen Clustern auf ein akzeptables Maß behandelt.

Zusätzlich wird eine einfache Lösung zur Intra-Cluster-Ressourcenzuweisung vorgestellt.

Realistische Szenarien von Fahrzeugen, die in einem Mobilfunksystem der fünften Generation simuliert werden, liefern Resultate, welche die Fähigkeit der vorgeschlagenen Systeme zur effektiven Anpassung an unterschiedliche KAF-Anforderungen belegen. Selbst bei hoher FZN-Verbindungsverzögerung bietet die langsame KZI-basierte zentrale FRP eine hohe FZF-Verbindungsqualität, welche die strengen Zuverlässigkeitsanforderungen erfüllt. Dies geht jedoch zu Lasten der Ressourcennutzung und somit der Paketverzögerungsgüte in FZF-Verbindungen. Dennoch zeigt ein semi-zentralisiertes Schema erhebliche Gewinne hinsichtlich der Paketverzögerung bei gleichzeitiger Aufrechterhaltung einer hohen FZF-Verbindungsqualität.

Abstract

Automated driving is emerging as a key trend and presents an opportunity for coordination among autonomous vehicles to improve road traffic flow, fuel efficiency and safety. In this regard, co-operative automated driving (CAD) is a means for vehicles to perform co-operative trajectory planning and enhance environment perception by leveraging their vehicle-to-vehicle (V2V) communication capabilities via co-located vehicle user-equipment (V-UE). Typical examples of CAD include platooning, co-operative lane change (CLC) and co-operative collision avoidance (CCA). The quality of service (QoS) requirements for CAD V2V communications span a wide range with a reliability of 90-99.999 % and an end-to-end packet delay of 3-100 ms. Here, the flexibility of QoS requirements reflects the performance tolerance depending on the use cases and different levels of automation (LoA) at the expense of suboptimal operations.

Sidelink (direct V2V link) radio resource management (RRM) plays a key role in meeting the CAD QoS requirements with QoS-aware radio resource scheduling (RRS). This requires tight interaction between CAD applications and the sidelink RRM. In addition, CAD necessitates V2V interactions and interference coordination among V-UEs across mobile network operators (MNOs), and vertical partners (e.g., automotive organizations) may wish to control sidelink radio resources. Therefore, the sidelink RRM needs to be abstracted or decoupled from the operator's network. To this end, the sidelink RRM functionality is moved to operator-independent cloud-based network entities that can be accessed by V-UEs across MNOs.

A cloud-enabled sidelink RRM framework is introduced which allows for placement of sidelink RRM functionality in a logically centralized cloud entity in a multi-

operator environment. By capturing the diverse QoS requirements of CAD multicast group communication using utility functions, a QoS-dependent utility-based multi-objective radio resource allocation optimization problem, which aims to maximize the aggregated utility, is described, wherein the resources refer to resource blocks (RBs) and modulation and coding schemes (MCSs). Moreover, a low-complexity heuristic centralized RRS solution is proposed and analyzed for cloud-based sidelink RRM with potentially high vehicle-to-network (V2N) link delay and a dynamic vehicular environment. Accounting for the challenges in acquiring channel state information (CSI), the dynamic RRS solution is enhanced to make use of slowly varying large-scale channel parameters (e.g., path loss) referred to as slow CSI.

An edge-cloud enabled semi-centralized sidelink RRM framework is proposed to bring the dynamic scheduler closer to the V-UEs by leveraging the vehicular edge computing (VEC) capabilities of the autonomous vehicles. To exploit the benefits of the framework, the resource allocation problem is split into three tasks with high modularity and tunability. These tasks comprise a V-UE-cluster formation, an inter-cluster RB-pool allocation and an intra-cluster resource allocation allowing for optimized sidelink resource scheduling, which may be performed on different time scales at different cloud entities (central and edge) with reduced dynamic scheduling delay. A graph-theoretic approach is presented with 1) a CAD vehicle group clustering as a clique partitioning problem (CPP), which enables the coordination at the control plane between transmitting vehicles to avoid the hidden node problem and the limitations imposed by the half-duplex constraint, and 2) a cluster-to-RB-pool allocation as a max-min fairness problem on a weighted resource-conflict graph, which tackles the trade-off between enhancing spectrum utilization by efficient reuse of sidelink resources among clusters and limiting the inter-cluster interference to an acceptable level. In addition, a simple low-complexity intra-cluster resource allocation solution is presented.

Simulation results in a realistic vehicular deployment in a fifth generation mobile networks (5G)-based simulation set-up show the ability of the proposed schemes to effectively adapt to different CAD requirements. Even in case of high V2N link delay, the slow CSI-based central RRS provides high V2V link quality in meeting the stringent reliability requirements. This, however, comes at the expense of resource utilization and hence the packet delay performance on V2V links. Yet, a semi-centralized scheme shows significant gains in terms of packet delay while maintaining a high V2V link quality.

Contents

List of Figures

List of Tables

Notation

Throughout the thesis boldface lower-case, boldface upper-case and calligraphic letters define vectors, matrices and sets, resp., where x_i stands for the i^{th} element of vector $\mathbf{x}$ and $Y_{i,j}$ refers to the entry in the i^{th} row of the j^{th} column of matrix $\mathbf{Y}$. Besides, $\mathbf{0}_{A\times B}$ and $\{Y_{i,j}\}$ denote the $(A \times B)$-dimensional zero matrix and the matrix of entries $\{Y_{i,j}\}$, respectively. Furthermore, $(\cdot)^{\mathrm{T}}$, $|\cdot|$, $\Pr(\cdot)$, $\log(\cdot)$, $\exp(\cdot)$, $\lfloor\cdot\rfloor$, $f'(\cdot)$ and $\mathbb{1}_{\mathcal{X}}(\cdot)$ denote transposition, cardinality, probability, natural logarithm, natural exponential, the floor function, derivative of function f and the indicator function on the set $\mathcal{X}$, respectively. Moreover, $\mathbb{N}$, $\mathbb{N}_0$ and $\mathbb{R}_0^+$ denote the sets of natural numbers and non-negative integer and real numbers, respectively. Finally, $\langle\mathbf{x}\rangle$ is a vector containing the elements of $\mathbf{x}$ in non-decreasing order and $\Omega_{a,b} = \{n \in \mathbb{N}_0 : a \leq n \leq a + b - 1\}$ is an index set of cardinality b defined for $a \in \mathbb{N}_0$, $b \in \mathbb{N}$.

Chapter 1

Introduction

1.1 Motivation

1.1.1 Co-operative Automated Driving

Automated driving is emerging as a key trend in the transportation sector. By the year 2045, as much as half of new vehicle sales and 40% of vehicle travel could be autonomous [1]. Freight trucks and long-haul buses are seen as candidates for early adoption of autonomous operation owing to their predictable routes and high labor costs. In addition, shared autonomous vehicles such as taxis are seen as attractive for reduced operating cost. Therefore, the autonomous trucks, buses and taxis are expected to be widely deployed in the next two decades [1].

Vehicles are increasingly being equipped with sensors, on-board data storage and computation resources to support different levels of automation (LoA). Each level reflects the degree of real-time operational (e.g., steering, accelerating, braking) and tactical (e.g., maneuver planning) functional capability of an automated driving system (ADS) to operate a vehicle in on-road traffic. The society of automotive engineers (SAE) International defines six LoA ranging from *no automation* (level 0) to *full automation* or fully self-driving vehicle (level 5) [2]. This draws a distinction between lower LoA and higher LoA based on whether a human driver or the ADS is primarily responsible for a dynamic driving task (DDT), which involves lateral

vehicle motion control (VMC) via steering, longitudinal VMC via acceleration and deceleration, and monitoring the driving environment via object and event detection and response (OEDR). The LoA are captured in Table 1.1. In case of no automation

Table 1.1: Levels of automation [2].

Level	Name	DDT Performer	
		VMC	OEDR
0	No Automation	Human	Human
1	Driver Assistance	Human and ADS	Human
2	Partial Automation	ADS	Human
3	Conditional Automation	ADS	ADS
4	High Automation	ADS	ADS
5	Full Automation	ADS	ADS

(level 0), the entire DDT is performed by the human driver. The lower two LoA (levels 1-2) refer to cases in which the human driver continues to perform a part of the DDT while the ADS is active. On the other hand, the higher three LoA (levels 3-5) refer to cases in which the ADS performs the entire DDT on a sustained basis while it is active [2], [3]. A narrative definition of LoA is provided in Appendix A.

The transportation sector has been facing challenges such as growing emission, increasing road situation complexity (e.g., road traffic congestion) and safety concerns due to, e.g., rapid urbanization, increase in number of vehicles on the road. To this end, the increasing adoption of autonomous vehicles presents an opportunity for coordination among autonomous vehicles to reduce greenhouse gas emission and to improve road traffic flow, fuel efficiency and safety. In this regard, co-operative automated driving (CAD) is a means for vehicles to perform co-operative trajectory planning and control as well as co-operative perception and road environment sensing by leveraging their communication capabilities [4]. CAD allows vehicles to drive closely for fuel saving thanks to reduced aerodynamic drag and for increased road utilization. In addition, it also allows vehicles to jointly perform maneuvers such

as lane change and collision avoidance in timely manner for smoother road traffic flow and increased safety. It is noted, 5G automotive association (5GAA) predicts all new autonomous vehicles to support connectivity from year 2026 [5], which can then be exploited to realize CAD. Typical use cases of CAD include

- platooning: Vehicles communicate to travel in a group with reduced inter-vehicular distance (IVD) as illustrated in Fig. 1.1. A smaller IVD reduces aero-

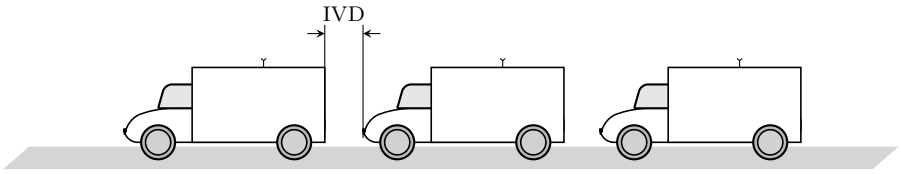


Figure 1.1: Illustration of high-density platooning of three vehicles.

dynamic drag experienced by the trailing vehicles, which reduces the overall fuel consumption (up to 20% depending on the IVD) and thus the operational cost for vehicle owners along with savings in greenhouse gas emissions [6]. Furthermore, the smaller IVD increases the road utilization reducing thereby the road congestion and improving the road traffic flow [7].

- co-operative lane change (CLC): Vehicles communicate to coordinate an autonomous lane change operation for one or more vehicles in a safe manner with reduced impact on the traffic flow [8], [9]. Fig. 1.2 illustrates a CLC operation

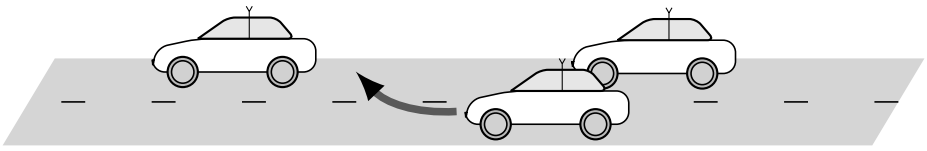


Figure 1.2: Illustration of CLC.

where three vehicles are coordinating to facilitate lane change of the vehicle on the left lane by creating safe lane changing gaps on the right lane. It is noted, up to 10% of road accidents are reported to have caused due to inappropriate lane changes in the U.S. [10]. In addition, an uncoordinated lane change op-

eration with delayed and increased safety gap creation adversely affects road traffic flow, which may cause traffic congestion.

- co-operative collision avoidance (CCA): When the vehicles are at risk of a collision, they communicate to coordinate their movement to avoid the collision and return to their original course when the risk is averted [11]. This significantly increases the road safety of the vehicles.

In CAD, the best control actions and trajectories are calculated for the involved vehicles to allow them to safely reach their destination (or target position) in the coordination area in finite time (e.g., within a few tens of seconds in case of CLC) while optimizing a performance objective such as fuel efficiency or quicker safety gap creation. Here, the control actions refer to real-time operational functions for VMC (e.g., steering, accelerating, braking). The computed control actions/trajectories must, at first, ensure safety for the involved vehicles so that they may never be steered to states from which future collisions are unavoidable. Secondly, the liveness (i.e., no permanent stop) must be guaranteed for all vehicles so that respective destinations are reached eventually without creating a traffic deadlock. That is, all involved vehicles must be allowed to both enter and exit the coordination area in finite time without coming to a permanent stop. In this regard, CAD can be formulated as a constrained optimization problem of finding the best control actions and/or trajectories for the involved vehicles that maximize a performance criterion (e.g., fuel efficiency) subject to safety and liveness requirements, and can be tackled by solving a finite-time optimal control problem at every time instant [4].

Vehicle-to-vehicle (V2V) communications facilitate the co-operative information exchange among autonomous vehicles that is required to formulate (based on, e.g., vehicle state, sensor data and trajectories) and solve (using, e.g., messages in a distributed iterative algorithm) the control problem. Therefore, the communica-

tion system forms a bottleneck in the coordination control since the impairments in communication may result in a potentially suboptimal solution. For example, a large packet delay[1], a low data rate (low trajectory resolution) and/or low reliability over a V2V communication link may lead to a larger IVD due to an increased IVD safety gap and slower maneuver execution. Consequently, in case of platooning, one may face a lower road utilization and a lower fuel efficiency gain [6]. In addition, higher LoA lead to stringent performance requirements in terms of packet delay, reliability and data rate due to safety reasons. Therefore, a control-aware communications system is necessary to carry out the coordination algorithm in CAD. To design the latter, CAD can be translated to quality-of-service (QoS) requirements in terms of data rate, reliability and packet delay requirements in the communication system. More precisely, the algorithm has to deal with a wide range of 90-99.999% reliability, end-to-end packet delays of 3 -100 ms and data rates of 0.012 -1000 Mbps [7], [12] depending on the use cases and different LoA. Here, the elasticity level of QoS requirements reflects the performance tolerance at the cost of suboptimal operations (w.r.t., e.g., fuel and/or road efficiency) [7] in the respective CAD operation and LoA. Table 1.2 captures maximum packet delay and reliability performance requirements for platooning and CLC for low and high LoA in fifth generation mobile networks (5G) according to the 3rd generation partnership project (3GPP) [12].

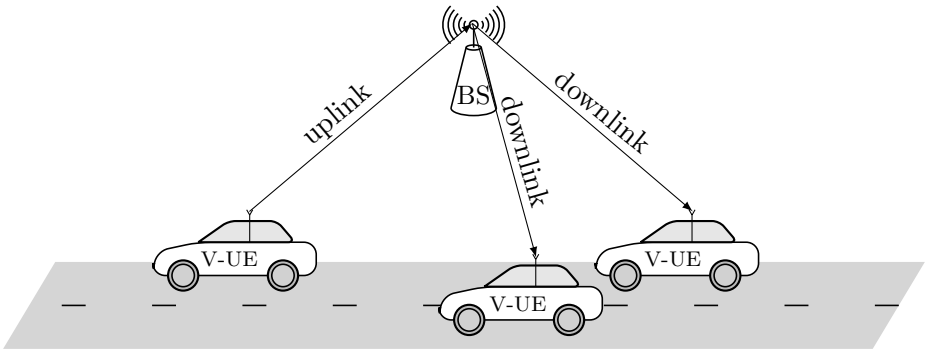
Table 1.2: 5G performance requirements for platooning and CLC.

use case	LoA	max. packet delay (ms)	reliability (%)
platooning	low (level 1/2)	25	90
	high (level 5)	10	99.99
CLC	low (level 1/2)	25	90
	high (level 5)	10	99.99

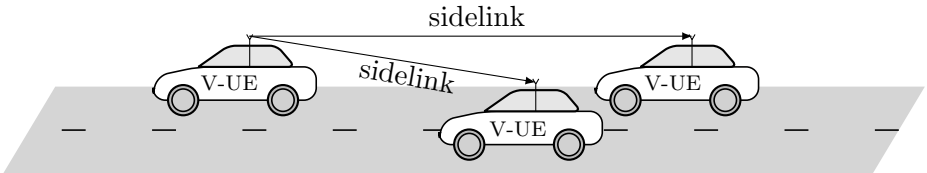
[1]Delay and latency are used interchangeably in this work.

1.1.2 Sidelink RRM for V2V Communication

To enable V2V communication between vehicles, each vehicle can be equipped with a communication module referred to as vehicle user-equipment (V-UE). V-UEs can communicate with each other either using an indirect V2V radio link via a network infrastructure node such as a base station (BS) or using a direct V2V radio link, known as *sidelink*, as illustrated in Fig. 1.3.



(a) Indirect V2V link via BS



(b) Direct V2V link (sidelink)

Figure 1.3: Illustration of indirect and direct V2V links.

In case of V2V communication using indirect V2V links, known as indirect V2V communication, the V2V data from a transmitter V-UE to a receiver V-UE traverses through a network infrastructure node using the vehicle-to-infrastructure (V2I) link between the transmitter V-UE and the infrastructure node, referred to as uplink, and the radio link between the network infrastructure node and the receiver V-UE, referred to as downlink. The indirect communication offers long-range V2V communication thanks to the involvement of one or more infrastructure nodes for relaying the V2V data. However, this increases the load on the infrastructure node as the

data traverses through it and hence it may act as a bottleneck for V2V communication resulting in increased packet delay. In addition, the infrastructure nodes are a single point of failure and may potentially disrupt the V2V communication in case of any malfunction. On the other hand, V2V communication using direct V2V links is known as direct or sidelink communication where the V2V data from the transmitter V-UE is sent directly over the sidelink to the receiver V-UE in close proximity without traversing through any network infrastructure node. Hence, it does not introduce any V2V data load on the infrastructure nodes and is not impacted by any bottleneck at the infrastructure node. Also, due to one-hop communication, it offers low latency V2V data exchange between V-UEs. Furthermore, since the receiver V-UE may be significantly nearer than the infrastructure node (e.g., BS) to the transmitter V-UE, the sidelink communication may also be more power-efficient as the transmitter V-UE may use less power to transmit over the sidelink to the receiver V-UE. It is noted that the communication range over a direct V2V radio link is limited and, hence, the sidelink V2V communication is suitable only for V2V data exchange among V-UEs that are in close proximity. This allows for spatial radio frequency resource reuse among distant V-UEs to increase spectral efficiency. However, an effective radio resource scheduling (RRS) is necessary to limit the interference at the receiving V-UEs to meet reliability requirements in sidelink communication while increasing spectral efficiency.

The vehicles involved in a CAD operation *are* in close proximity. Therefore, considering the aforementioned advantages over indirect communication, this work employs sidelink V2V communication for CAD V2V communication. Here, sidelink radio resource management (RRM) plays a key role in meeting the wide range of CAD requirements on sidelink V2V links. The sidelink RRM involves management of radio spectrum, co-channel interference and transmission parameters such as the modulation and coding scheme (MCS) and the transmission power for sidelink

communication. By means of RRS, a V-UE can be allocated with radio resources including radio spectrum resources (or resource blocks (RBs)), transmission power and MCS for a sidelink transmission. This allows for controlling the data rate, interference management as well as link adaptation. Hence, this work focuses on sidelink RRM to support CAD V2V communications.

Depending on whether a network infrastructure node assistance is available in radio resource allocation or not, currently there are two major sidelink resource allocation modes, namely, network-assisted and distributed modes. The network-assisted mode refers to the case where the infrastructure node, e.g., a BS assists in sidelink resource allocation. Here, typically, a BS schedules sidelink radio resources to be used by the transmitting V-UE (V-UE 2 and V-UE 5 in Fig. 1.4) and indicates the allocated resources, referred to as *schedule grant*, to the V-UEs as illustrated in Fig. 1.4. LTE-V Mode-3 [13] and new radio (NR) vehicle-to-everything (V2X) Mode-1 [14]

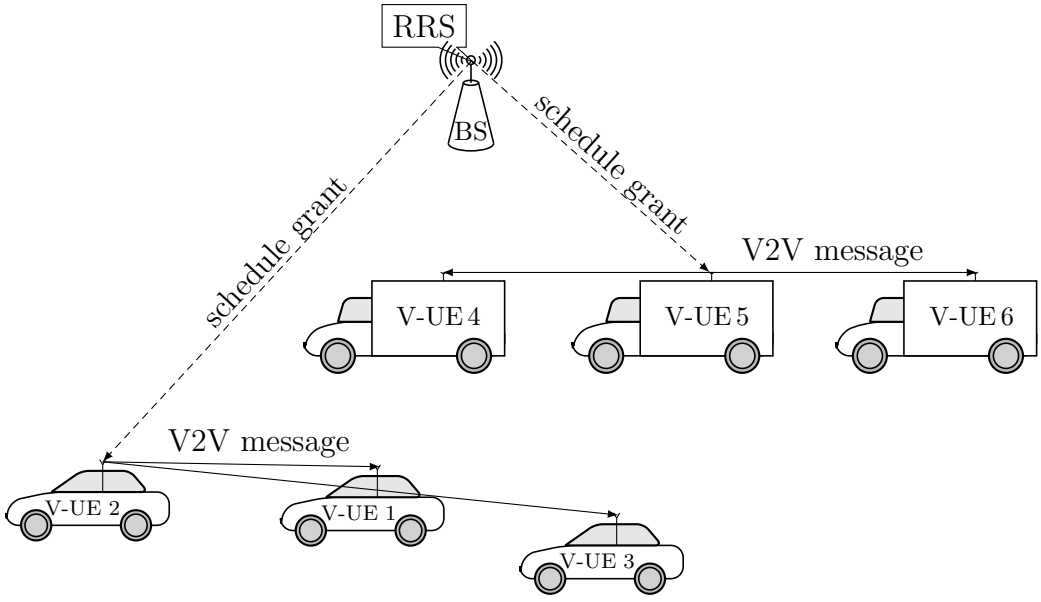


Figure 1.4: Illustration of network-assisted sidelink radio resource allocation.

are notable network-assisted resource allocation-based V2V communication standards from 3GPP where eNodeB (eNB) (a 3GPP-complaint implementation of a long-term-evolution (LTE) BS) and gNodeB (gNB) (a 3GPP-complaint implemen-

tation of a 5G NR BS), resp., peform sidelink RRS. This mode offers a BS a high control over radio resource allocation and hence on data rate, link adaptation and interference management. Therefore, it has the potential to provide better sidelink performance (e.g., high reliability, low latency). However, it faces the challenge of higher signaling overhead on uplink and downlink transmissions because of control plane signaling related to RRM, referred to as radio resource control (RRC) signaling, between BS and V-UEs (e.g., schedule grant, channel measurements). Moreover, it requires a single mobile network operator (MNO) to manage the radio resources. Furthermore, vehicles cannot be supported in case of radio coverage void. On the other hand, in the distributed mode, V-UEs select radio resources for sidelink transmission in a distributed manner by employing a random selection or channel sensing-based schemes. A few notable V2V communication standards based on distributed resource selection (e.g., channel sensing-based) are IEEE 802.11p from Institute of Electrical and Electronics Engineers (IEEE) [15], LTE-V Mode-4 [16] and NR V2X Mode-2 [14] from 3GPP. Fig. 1.5 illustrates a distributed resource allo-

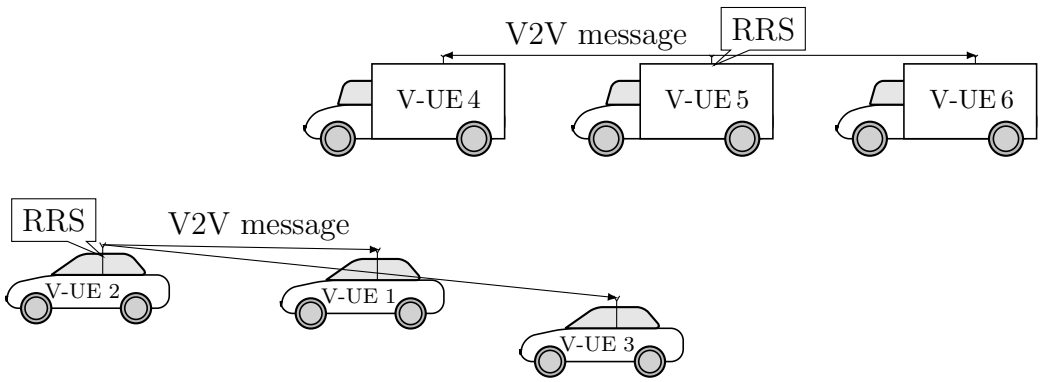


Figure 1.5: Illustration of distributed sidelink radio resource allocation.

cation scenario where each transmitter V-UE (V-UE 2 and V-UE 5) performs RRS, i.e., selects radio resources autonomously for its sidelink transmission. The random selection, where V-UEs randomly select a resource for sidelink transmission, is simpler to implement, but offers low reliability due to its susceptibility to high packet

collision. The channel sensing-based schemes offer higher reliability than the random selection schemes since the V-UEs can avoid accessing those radio spectrum resources in which it expects interference to its transmission based on channel sensing measurements. However, with an only local view of the interference situation, the sensing-based schemes can suffer from the hidden node problem and half-duplex limitations. Also, it can suffer from the exposed node problem and hence may result in inefficient resource utilization. Consequently, while the fully distributed mode may not require any infrastructure (there is no cost on signaling over uplink/dowlink and on infrastructure deployment) and is not MNO dependent, the aforementioned drawbacks limit its ability to provide higher reliability for sidelink V2V communication.

QoS requirements of CAD applications span a wide range depending on the level of CAD operation efficiency (w.r.t., e.g., fuel efficiency, road efficiency), level of criticality of CAD operation (w.r.t., e.g., collision risk) and the operating LoA as described in Sect. 1.1.1. It is noted that the CAD applications demand high sidelink reliability, in particular for higher CAD operation efficiency (e.g., high fuel saving in platooning with low IVD), higher LoA and for critical use cases such as CCA. Therefore, considering the potential of network-assisted mode over distributed mode in offering high reliability for sidelink V2V communication, this work deals with network-assisted resource selection. Since the QoS performance on sidelinks directly impacts the CAD operation efficiency (e.g., fuel saving) and the LoA, sidelink radio resource allocation strategies are needed that take into account the diverse elasticity level of QoS requirements to support high CAD operation efficiency. In this regard, the sidelink RRM functionality of the network must be aware of the CAD application requirements and the involved V-UEs in CAD operation. Here, an efficient means is required which captures the elasticity level of the QoS requirements of the corresponding CAD application. In addition, it is essential for an application

to adapt the CAD operation (e.g., to increase IVD in platooning) by monitoring changes in the sidelink performance (e.g., increase in latency). Hence, there is a need for interaction between the sidelink RRM function and the CAD application.

CAD necessitates V2V interactions and interference coordination among vehicles across a given MNO. In a multi-operator environment, V-UEs may belong to different MNOs, where the sidelink transmission of a V-UE is controlled by the subscribed MNO as illustrated in Fig. 1.6. Consequently, V-UEs involved in a CAD operation

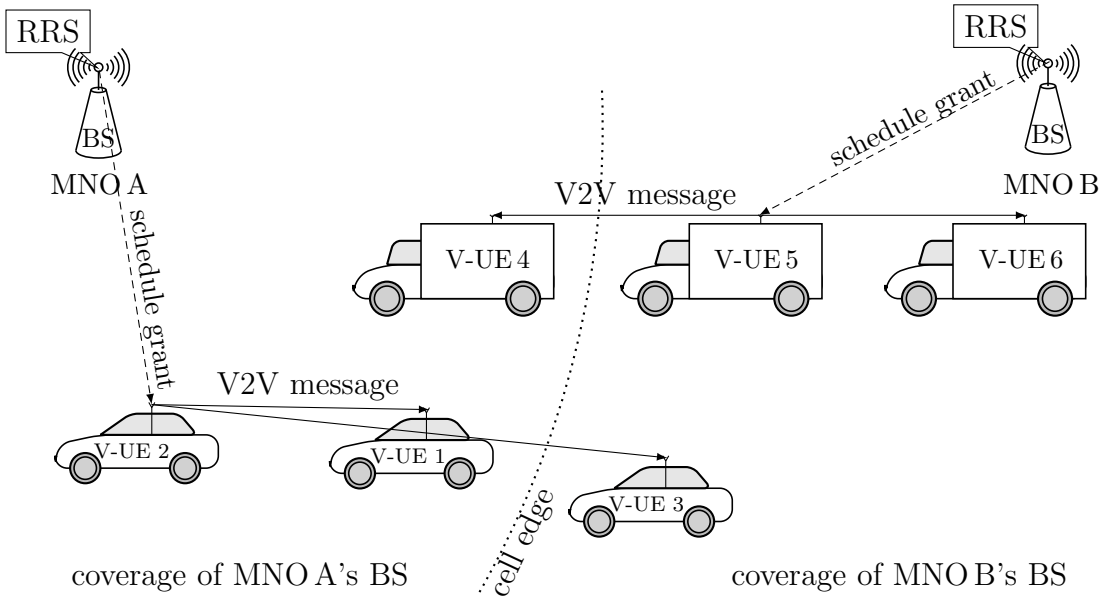


Figure 1.6: Illustration of network-assisted mode of sidelink V2V communication in a multi-operator scenario.

such as CLC, sensor sharing and safety critical CCA may belong to different MNOs. Therefore, V2V interactions and interference coordination must be ensured among V-UEs subscribed to different MNOs. In addition, sidelink RRM may need to be abstracted or possibly decoupled from the operators' network to allow sidelink radio resource control by vertical partners (e.g., car manufacturers) and applications.

Recently, radio access network (RAN) virtualization is gaining momentum where the RAN aspects/functions (e.g., baseband-processing) are being decoupled from the BS hardware and are being moved to a cloud computing platform with virtualized

computing infrastructure constituted from commodity hardware. This allows for programmability via an application programming interface (API) and for flexibility in deployment and scalability because of on-demand computing resource provisioning, and it can be cost effective. Virtualization of RAN computing resources permits a dynamic reassignment of processing resources to a BS, whereas spectrum virtualization enables multiple MNOs to share the same spectrum and allows for efficient spectrum utilization [17]. This RAN virtualization trend can be extended and exploited for sidelink RRM as moving the sidelink RRM function from the operator's BS to a network entity, a cloud server being accessible by different operators, can be significantly beneficial. This has the potential to allow for sidelink radio resource allocation with interference coordination across MNOs, and allows for vertical partners to control the sidelink radio resources. Also, since V2X applications are increasingly controlled by cloudified V2X application-specific servers (e.g., platooning servers), the placement of sidelink RRM function at the cloud can enable interaction with CAD applications. Therefore, this work aims to provide a cloud-enabled sidelink RRM framework, where RRM related tasks can be performed at cloud entities instead of MNOs' BSs.

The placement of a radio resource scheduler of sidelink RRM in a logically centralized cloud may allow for a global RRS across MNOs and BSs. Considering a shared radio spectrum band (e.g., the intelligent transportation system (ITS) band) across MNOs, such a global RRS avoids radio spectrum fragmentation, which otherwise is needed to avoid inter-cell interference and interference across MNOs. Hence, the RRS can flexibly allocate radio resources on-demand (i.e., when needed by the CAD application) with high radio resource utilization. However, this increases the control plane delay, i.e., RRC signaling delay because of the larger communication delay over vehicle-to-network (V2N) links between the scheduler placed in the remote cloud and the V-UEs due to high transport network delays, which depend on

the deployed backhaul topologies and technologies. Such high RRC signaling delay over V2N links increases scheduling delay and hence the latency which could potentially lead to sub-optimal solutions of the CAD control problem, which, in turn, may be critical for emergency maneuvers such as CCA. Due to a large degree of centralization, the centralized approach to the CAD control problem also suffers from an increased signaling overhead over uplink and downlink and the risk of outdated RRS-relevant channel measurements in dynamic vehicular environments, which is not desirable. Therefore, there is a need for RRS strategies that are robust against the large V2N link delay, referred to as V2N delay for short henceforth, and that reduce signaling overhead over uplink and downlink.

Since the RRC signaling delay between the scheduler and V-UEs is limited by the communication delay between the cloud entity and the V-UEs, edge computing [18] can be exploited to reduce the RRC signaling delay. In the edge computing paradigm, the computing resources are made available at the edge of the network near to or co-located with the end devices [19]. Hence, by placing the radio resource scheduler at the edge cloud, the RRC signaling delay can be significantly reduced. However, unlike the logically centralized cloud (or remote cloud), the edge cloud coverage may be limited (i.e., an edge cloud may have connectivity to only a subset of V-UEs that are in close proximity). Consequently, the scheduler placed at the edge cloud entity may only be able to schedule a subset of V-UEs. Therefore, interference management across the coverage areas of multiple radio resource schedulers that are placed in different edge cloud entities must be ensured while keeping the radio resource utilization high. Hence, this work aims to introduce an edge cloud-enabled sidelink RRM framework and RRM strategies which support sidelink RRS with reduced scheduling delay, controllable interference levels and reduced V2N signaling overhead on uplink and downlink to support low latency and high reliability sidelink CAD V2V communication.

1.2 State of the Art

1.2.1 Sidelink Radio Resource Allocation for V2V Communications

Aiming to reduce signaling overhead over V2I links (i.e., uplink/downlink) and/or scheduling delay in sidelink RRS as well as to make the RRS robust against large RRC signaling delay over V2I links, a number of strategies [20]–[30] have been proposed recently for V2V communications. In [20], two interference graph-based resource-sharing schemes are presented to maximize the network sum rate with reduced complexity. Here, the interference-aware approach gathers full channel state information (CSI)[2] for optimal interference management, whereas the interference-classified one uses only the relative geographic locations of the V-UEs to reduce signaling overhead, however, at the cost of sub-optimal operation. Furthermore, a location-based scheduling framework is proposed in [21] providing an optimal resource-reuse range to ensure constraints on communication range and reliability. Reducing the signaling overhead by considering only slowly varying large-scale channel parameters like path loss and shadowing, termed slow CSI henceforth, [22] attempts to maximize the V2I link throughput while guaranteeing a certain reliability on V2V links allowing spatial reuse of resources among V2I and V2V users. Transforming the latency and reliability requirements into optimization constraints that are computable using only the slow CSI, [23] proposes an RRM scheme that aims to maximize the sum rate over V2I links while satisfying delay and reliability requirements of V2V communications. [24] presents a low-complexity clustering-based resource allocation for strict latency and reliability constraints where the V-UEs that cause mutual interference are grouped into fixed number of clusters and the RBs are

[2]Full CSI and instantaneous CSI are interchangeably used in this work.

reused across clusters.

Dividing the cell sector into disjoint zones and providing each zone with a dedicated RB-pool, [25] allocates an exclusive RB from the RB-pool to a V-UE as it enters a new zone to reduce signaling overhead and complexity. On the other hand, [26] allocates orthogonal RB-pools proportional to the zone demand and performs scheduling per zone while considering latency requirements. Further reducing the signaling overhead, [27] proposes a zone-based hybrid scheduling mechanism with a two-period solution. Here, the BS schedules in the first period to assure link reliability requirement and the V-UEs autonomously select the resources based on channel sensing in the second period. Presenting a zone-based V-UE clustering and an intra-cluster vehicle level scheduling to reduce scheduling delay, [28] implements a round-robin scheduler. Furthermore, clustering spatially far apart V-UEs at first, [29] performs a matching game-based intra-cluster resource selection for improved link quality, whereas [30] aims to minimize the transmit power at the V-UE in the assigned RBs while meeting latency and reliability requirements.

When RRM functionalities are placed at a central cloud entity, centralized schemes [20]–[25] can reduce the signaling overhead over V2N links by acquiring slow CSI or location, but may introduce larger scheduling delay which could limit their ability to support CAD. Also, they do not treat small-scale channel variations in dynamic scheduling. On the other hand, the semi-centralized schemes in [26]–[30] allow for decoupling the scheduling. The schemes can be centrally cloud-assisted as they allow a two-stage resource allocation (with different time scales and at different cloud entities). The latter exploits the information which of the RRC signaling delays over V2N links can be eliminated in a dynamic scheduling by placing the scheduler at an edge cloud that is close to or co-located with the vehicle. However, these schemes do not consider V2V control plane reachability during V-UE clustering, which is crucial for reliable transmission coordination (joint scheduling).

Moreover, the aforementioned schemes do not flexibly adapt to multicast QoS requirements of CAD.

1.2.2 Edge Cloud Computing Framework for Vehicular Scenarios

A vehicular edge computing (VEC) architecture is proposed in [31] to utilize vehicles as the infrastructure for communication and computation. Here, vehicles collaboratively carry out communication (e.g., relaying packet exploiting multi-hop characteristic and moving features) and computation utilizing individual communication and computation resources of each vehicle. [32] proposes an autonomous vehicular edge framework specifying the workflow for autonomous organization of a vehicular cloud to manage the idle computing resources in a decentralized manner to increase the computational capabilities of vehicles. While [31], [32] turn each vehicle into a node of a distributed cloud platform, [33] proposes to place edge nodes on only certain vehicles in the VEC platform. Here, each vehicular edge node act as a local cloud server as well as a wireless access point and utilizes the mobility to provide cost-effective on-demand computing and communication capability.

A multi-access edge computing (MEC) paradigm is considered in [34]–[36] where the edge nodes, MEC servers, are placed at BSs and roadside units (RSUs). Here, [36] allows vehicles to offload computational task to the MEC servers, whereas [34], [35] utilitze MEC capabilities for assisted driving and infotainment by providing different services like high-definition real-time maps as well as real-time traffic monitoring and alerts. Furthermore, to enable cloud-based automotive services to follow vehicles over the mobile network in an MEC paradigm, [37] proposes a follow-me edge-cloud (FMeC) framework ensuring connection to the optimal edge via service migrations.

A hybrid vehicular edge cloud (HVC) is presented in [38], [39] which enables an efficient sharing of all accessible computing resources from VEC nodes, MEC nodes and the central cloud by using multi-access networks. [40] proposes a cloud-based MEC off-loading framework where MEC servers are interconnected by a wireless backhaul to provide a virtual computation resource pool along with vehicular computation and communication resources. Moreover, the approach in [41] includes both a logically centralized cloud as well as an edge cloud.

1.3 Contribution and Organization

This work makes the following contribution to sidelink RRM for V2V communications in CAD use cases:

- A cloud-enabled sidelink RRM framework is introduced for QoS-aware RRS at a logically centralized cloud server in an operator independent manner to enable sidelink V2V interactions for CAD and interference coordination among V-UEs in a multi-operator environment. The framework is extended by leveraging the VEC capabilities of autonomous vehicles to bring the dynamic scheduler closer to the V-UEs for low latency RRS. In this regard, an edge-cloud enabled semi-centralized sidelink RRM framework is proposed which allows for the RRM task to be split into long-term and dynamic tasks that can be performed in centralized cloud server and VEC nodes, resp., at different time scales to selectively achieve different RRM objectives (e.g., low signaling overhead over V2N links, high reliability and/or low latency on V2V links, etc.) and to effectively serve different CAD use cases.

- By analyzing the diverse degrees of QoS performance tolerance of different CAD applications, however, at the cost of suboptimal CAD operations (w.r.t., e.g., fuel or road efficiency), CAD QoS requirements have to be elastic or

soft real-time and hence are captured using utility functions. A novel QoS-dependent utility-based multi-objective sidelink radio resource allocation optimization problem is formulated for CAD multicast group communication, wherein the resources refer to RBs and MCSs. This aims to maximize the aggregated utility (hence, the CAD operation efficiency) across multicast groups with reliability constraints and half-duplex limitation.

- A low-complexity algorithmic centralized sidelink RRS solution is proposed for the CAD radio resource allocation problem and is analyzed for RRS in the cloud-enabled sidelink RRM framework with potentially large V2N delay and a dynamic vehicular environment. To allow for robust resource allocation against large V2N delay and to reduce signaling overhead over V2N links, a RRS scheme that makes use of slowly varying large-scale channel parameters (e.g., path loss) is presented.

- To exploit the edge-cloud enabled sidelink RRM framework to reduce scheduling delay and V2N signaling overhead, a semi-centralized RRS scheme is proposed where the resource allocation problem is split into three tasks with high modularity and tunability. The three tasks comprise a V-UE-cluster formation, an inter-cluster RB-pool allocation and an intra-cluster resource allocation allowing for optimized sidelink resource scheduling, which may be performed on different time scales at different cloud entities (central and edge) with reduced dynamic scheduling delay. A graph-theoretic approach is presented with

    1. a CAD vehicle group clustering as a clique partitioning problem (CPP) which enables the coordination at the control plane between transmitting vehicles to avoid the hidden node problem and the limitations imposed by the half-duplex constraint and

2. a cluster-to-RB-pool allocation as a max-min fairness problem on a weighted resource-conflict graph which tackles the trade-off between enhancing spectrum utilization by efficient reuse of sidelink resources among clusters and limiting the inter-cluster interference to an acceptable level.

In addition, a simpler low-complexity intra-cluster resource allocation solution is presented.

The thesis is organized into chapters where the content is based on the publications [42]–[51] and the chapters are organized as follows.

Chapter 1: CAD use cases and a wide range of QoS requirements for CAD V2V communication with different elasticity levels are discussed. Sidelink communication that enables communication between V-UEs in close proximity over a direct V2V link (sidelink) is introduced. Considering the importance of sidelink RRM in meeting the CAD requirements, different sidelink resource allocation modes are presented and a need for QoS-aware and MNO-independent network-assisted RRM strategies is identified. The potential benefits and challenges in placing the sidelink RRS function in centralized cloud and edge cloud entities are discussed. In addition, the state of the art in sidelink RRM schemes and edge computing for vehicular scenarios is presented.

Chapter 2: The vehicular communication system considered in the thesis is described. This includes description of the involved network nodes and the associated communication links, CAD vehicular groups, packet generation at CAD application and packet queuing at V-UEs, and considered radio resources that are available for allocation. Furthermore, the considered models for sidelink channel and channel quality together with QoS parameters (reliability, data rate and delay) and QoS requirements are presented.

Chapter 3: A cloud-enabled sidelink RRM framework is introduced which allows for placement of sidelink RRM function in a logically centralized cloud entity for sidelink RRM abstraction from MNOs' networks and to allow interaction between the sidelink RRM and V2X applications. Then, by capturing the diverse QoS requirements of CAD at multicast groups using utility functions, a resource allocation problem that aims to maximize the aggregated QoS-dependent utility is described and an iterative low-complexity solution is proposed and analyzed for cloud-based sidelink RRM. Accounting for the challenges in acquiring CSI, a low-complexity scheduling scheme is presented that makes use of slow CSI.

Chapter 4: An edge cloud-enabled semi-centralized sidelink RRM framework is presented where the sidelink RRM task is split into long-term RRM and dynamic RRS tasks to reduce signaling overhead over V2N links and scheduling delay over V2V links. Then, for RRS under this framework, the resource allocation problem is split into three tasks, namely a V-UE cluster formation, an inter-cluster RB-pool allocation and an intra-cluster resource allocation allowing for optimized sidelink resource scheduling with reduced dynamic scheduling delay. Subsequently, a graph-theoretic approach is presented for V-UE clustering and inter-cluster RB-pool allocation. Furthermore, a simple intra-cluster RRS scheme is discussed.

Chapter 5: Simulating a realistic dynamic vehicular topology for a platooning use case with two LoA (low and high), the performance of all the proposed RRS schemes is evaluated on a common simulation platform based on 5G. At first, the 5G-based simulation set-up, performance metrics and simulated schemes are described. Then, the simulation results are presented and discussed for the proposed schemes.

Chapter 6: Finally, conclusions of the work are drawn. After a summary of the work and a discussion of the results, an outlook for potential extensions of the work is presented.

Chapter 2

System Description

This chapter describes the vehicular communication system considered in this work for designing robust sidelink RRS strategies to support CAD applications. At first, the considered network nodes and the associated communication links in the vehicular network are presented in Sect. 2.1. Since CAD is a group-based service, CAD vehicular groups are defined in Sect. 2.2. Sect. 2.3 describes packet generation by CAD applications and packet queuing at V-UEs, where each packet is assumed to include CAD information. Sect. 2.4 describes the considered radio resources that are available for allocation, and Sect. 2.5 presents sidelink channel model and defines sidelink channel quality for group communication. Subsequently, QoS parameters are defined in Sect. 2.6 and the QoS model is presented in Sect. 2.7.

2.1 Vehicular Network

Fig. 2.1 illustrates the considered vehicular network. The network consists of autonomous vehicles equipped with V-UEs, a cloud server and BSs as described in the following.

- **Vehicles**: Consider road-bound autonomous vehicles that can support different LoA. Each vehicle has on-board data storage and computational resources (e.g., non-volatile memory, processor). To enable radio communication capability, each vehicle is equipped with a communication module referred to

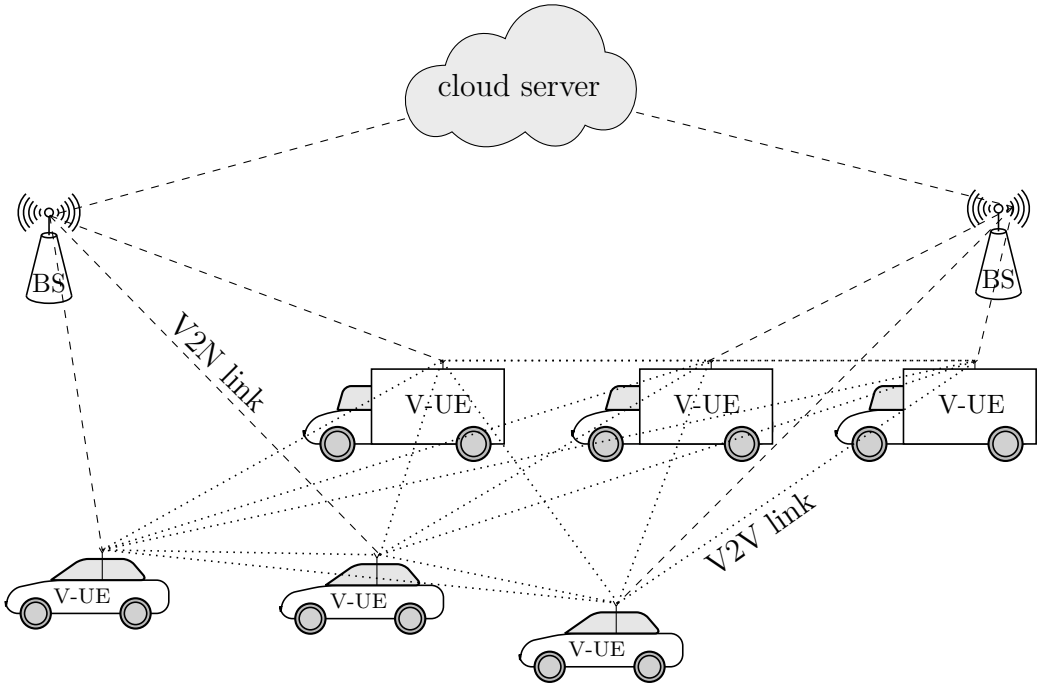


Figure 2.1: Illustration of cloud-assisted vehicular network.

as V-UE which includes a single omni-directional antenna and orthogonal frequency-division multiplexing (OFDM) transmitter and a receiver. We let $D \in \mathbb{N}$ be the number of vehicles and, hence, the number of V-UEs on a given road topology.

- **Cloud Server**: A cloud server with large storage and computational resources is considered. Here, it is a centralized logical entity whose implementation may include geographically distributed and remotely placed multiple physical servers as well as data centers. The cloud server is remotely accessible via the Internet.

- **Base Stations**: BSs are equipped with transceivers to provide radio coverage to V-UEs for uplink and downlink communication. Acting as a gate-way to the Internet, BSs enable V-UEs to connect to the remotely placed cloud server via the Internet.

V-UEs use V2V links to exchange both payload information (e.g., CAD data) and control plane information (e.g., RRC signalling) while utilizing V2N links to exchange control plane information (e.g., RRC signalling) with the cloud server via BSs as described below.

- **V2V Links**: Using transceivers, V-UEs can communicate with each other over direct V2V radio links known as *sidelink*. For CAD, V-UEs exchange data as data plane (or payload) as well as control plane information over direct V2V links. Let $\mathrm{L}_{d,\tilde{d}}$ denote the V2V link between the d^{th} transmitter and the $\tilde{d}^{\text{th}}$ receiver V-UEs with $\left(d, \tilde{d}\right) \in \Omega^2_{1,D},\ \tilde{d} \neq d$. Then, $\mathcal{L} = \{\mathrm{L}_{d,\tilde{d}} : (d, \tilde{d}) \in \Omega^2_{1,D}, \tilde{d} \neq d\}$ is the set of all direct V2V links.
- **V2N Links**: V-UEs use V2N links to exchange RRC information with the remote cloud server via BSs over the transport network. V2N links are assumed to be always available for each V-UE and to introduce a fixed delay of T_{V2N}.

It is noted that, in this work, V2N links do not carry any payload (i.e., CAD information) and are utilized to exchange only sidelink RRM related control plane information with the cloud server. Furthermore, this work does not optimize QoS performance over V2N links, but assumes V2N links to be always available (lossless) with fixed latency performance.

2.2 Vehicular Groups

CAD is a group-based service which involves a group of V-UEs to perform V2V communication within the group to exchange co-operative information.

In this work, a set of V-UEs which are participating (or co-operating) to carry out a CAD operation is referred to as a *CAD group*. Fig. 2.2 illustrates two CAD groups, where CAD group 1 consists of V-UEs $1, 2$ and 3 performing, e.g., CLC, and CAD group 2 consists of V-UEs $4, 5$ and 6 performing, e.g., platooning. Let D

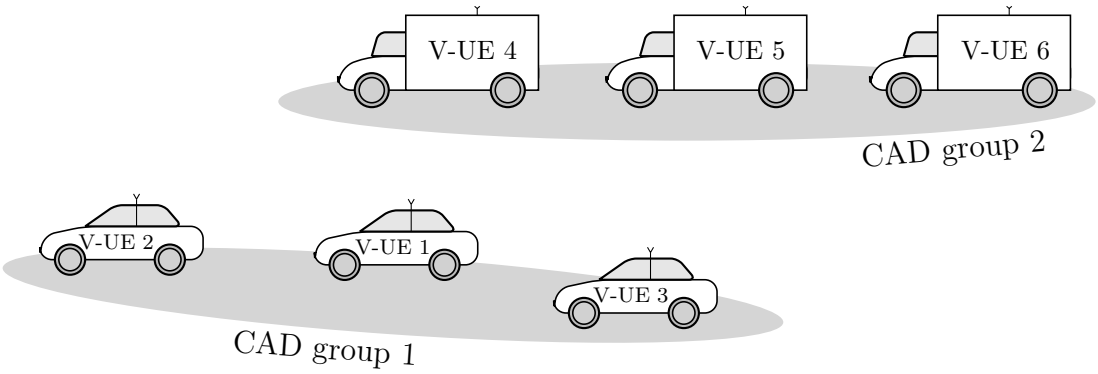
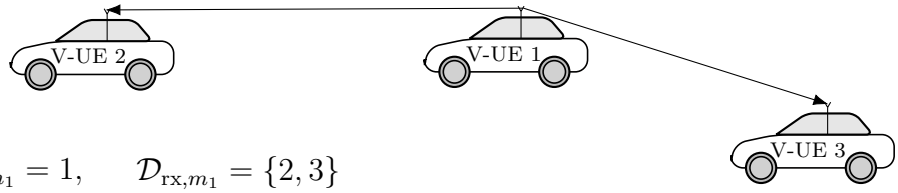


Figure 2.2: Illustration of CAD groups: CAD group 1 and CAD group 2 of three V-UEs performing e.g. CLC and platooning, respectively.

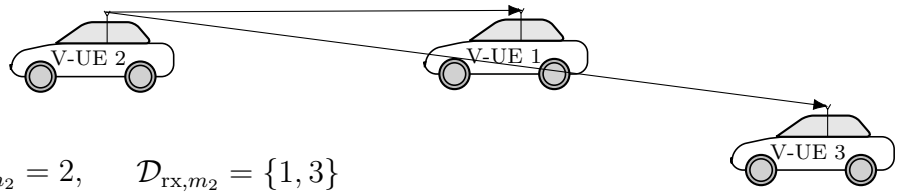
V-UEs on the road topology be divided into $X \in \mathbb{N}$ CAD groups.

In each CAD group, the participating V-UEs perform multicast group V2V transmissions (groupcast) to other V-UEs within the CAD group. In the system, to uniquely identify a single transmitting and a set of receiving V-UEs forming a pair for a multicast transmission, *multicast groups* are defined. Therefore, a multicast group consists of a transmitter V-UE and one or more receiver V-UEs. Each CAD group consists of multiple multicast groups. Let M be the number of multicast groups across CAD groups in the system and let $\mathcal{M}_{\text{CAD},x} \subseteq \Omega_{1,M}$ be the set of multicast groups within the x^{th} CAD group with $x \in \Omega_{1,X}$. Denoting the index set of all the member V-UEs as $\mathcal{D}_m \subseteq \Omega_{1,D}$, let $d_{\text{tx},m} \in \mathcal{D}_m$ and $\mathcal{D}_{\text{rx},m} \subset \mathcal{D}_m$ with $d_{\text{tx},m} \notin \mathcal{D}_{\text{rx},m}$ be the transmitter V-UE index and the set of recipient V-UE indices, resp., for the m^{th} multicast group with $m \in \Omega_{1,M}$. Fig. 2.3 illustrates multicast groups withing a CAD group. In this example, the x^{th} CAD group consists of V-UEs 1,2 and 3, where each V-UE performs multicast transmissions to all the other V-UEs in the CAD group to support CAD operation. Hence, there exist three multicast groups in this CAD group. Here, the first (${m_1}^{\text{th}}$) multicast group has V-UE 1 as the transmitter V-UE and V-UEs 2 and 3 as receiver V-UEs. The second (${m_2}^{\text{th}}$) multicast group has V-UE 2 as transmitter V-UE and V-UEs 1 and 3 as receiver V-UEs. Finally,



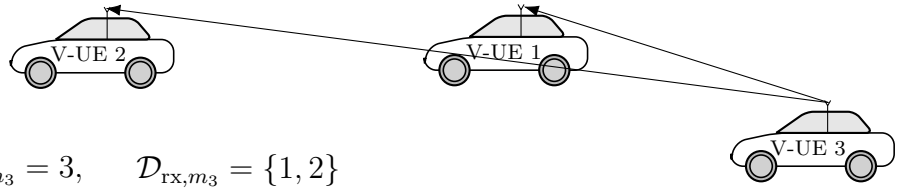
$d_{\text{tx},m_1} = 1, \quad \mathcal{D}_{\text{rx},m_1} = \{2, 3\}$

(a) multicast group of index m_1



$d_{\text{tx},m_2} = 2, \quad \mathcal{D}_{\text{rx},m_2} = \{1, 3\}$

(b) multicast group of index $m_2 \neq m_1$



$d_{\text{tx},m_3} = 3, \quad \mathcal{D}_{\text{rx},m_3} = \{1, 2\}$

(c) multicast group of index m_3 with $m_3 \neq m_2$ and $m_3 \neq m_1$

Figure 2.3: Illustration of multicast groups in the x^{th} CAD group involving V-UEs 1,2 and 3 and three multicast groups of indices m_1, m_2 and m_3, i.e., $\mathcal{M}_{\text{CAD},x} = \{m_1, m_2, m_3\}$.

the third ($m_3{}^{\text{th}}$) multicast group has V-UE 3 as transmitter V-UE and V-UEs 1 and 2 as receiver V-UEs.

In this work, the information on the CAD groups, i.e., the number of CAD groups and V-UEs within each CAD group, is considered to be provided by the V2X application layer. Besides, each V-UE in a CAD group is assumed to perform multicast transmission to support CAD. Therefore, the number of multicast groups the same as the number of V-UEs, i.e., $M = D$. In addition, a V-UE is assumed to multicast (or groupcast) to all the other V-UEs in the CAD group, and hence $\mathcal{D}_{\text{rx},m} = \mathcal{D}_m \setminus \{d_{\text{tx},m}\}$. Furthermore, a V-UE is assumed to be involved in at most one CAD operation and is therefore part of at most one CAD group. Finally, $\mu_d \in \Omega_{1,M}$ with $d \in \Omega_{1,D}$ denotes the index of the multicast group in which the d^{th} V-UE is the transmitter.

2.3 Packet Generation and Queueing

A CAD application at the transmitter of a multicast group is considered to generate packets that are assumed to include CAD information. Let $B_m \in \mathbb{N}$ be the packet size in bits of each packet that is generated by m^{th} multicast group. Depending on the nature of CAD information, a multicast group generates packets either periodically and non-periodically as described in the following.

- **Periodic data traffic (PDT)**: A CAD application may generate CAD messages periodically in case the CAD operation requires information exchange at regular intervals. For example, in CCA, vehicles may exchange their velocity, geographical location and/or trajectories at regular intervals to detect and avoid potential collisions. It is assumed that the CAD application generates packets at the m^{th} transmitter multicast group where the CAD operation requires periodic information exchange at regular intervals. Here, the CAD application is assumed to generate one packet at a given interval with $T_{\text{pkt},m} \in \mathbb{N}$ denoting the time interval in slots.

- **Non-periodic data traffic (NPDT)**: A CAD application may generate CAD messages non-periodically where the message generation is triggered, for example, based on observed events such as collisions, hazardous road situations, etc. In such a scenario, the arrival rate of packets by the CAD application at the m^{th} multicast group is considered to follow a Bernoulli process with a mean packet arrival rate $1/T_{\text{pkt},m}$.

For simplicity, mixed data traffic, which consists of both periodically and non-periodically generated packets, is not considered.

The transmitting V-UE of each multicast group is considered to maintain a queue that is implemented as a first-in first-out (FIFO) buffer of finite length Δ_m in packets.

Here, Δ_m is assumed to be N. Let $q_{m,\delta,n}$ denote the number of untransmitted bits of the packet enqueued at the beginning of time slot $n-\delta$, $\delta \in \mathbb{N}_0$, for the m^{th} multicast group. Then, the number of bits in the queue at the beginning of time slot n for the m^{th} multicast group can be written as

$$Q_{m,n} = \sum_{\delta=0}^{\Delta_m - 1} q_{m,\delta,n}. \tag{2.1}$$

Since the packets are not expected to be served in the same time slot as they have arrived [52], the queuing delay of the packet is set to one when it is enqueued in the buffer.

2.4 Sidelink Radio Resources

The system assumes sidelink orthogonal frequency-division multiple access (OFDMA) in a dedicated radio spectrum band. Each radio frame is divided into slots of duration T_{sl} and the system bandwidth is partitioned into $K \in \mathbb{N}$ orthogonal RBs. RBs are the smallest unit of frequency resources that are available for allocation at every time slot $n = 1, 2, \ldots, N$. Here, each RB consists of a fixed number of subcarriers and symbols per subcarrier as shown in Fig. 2.4. Consider K_{RE} to be the

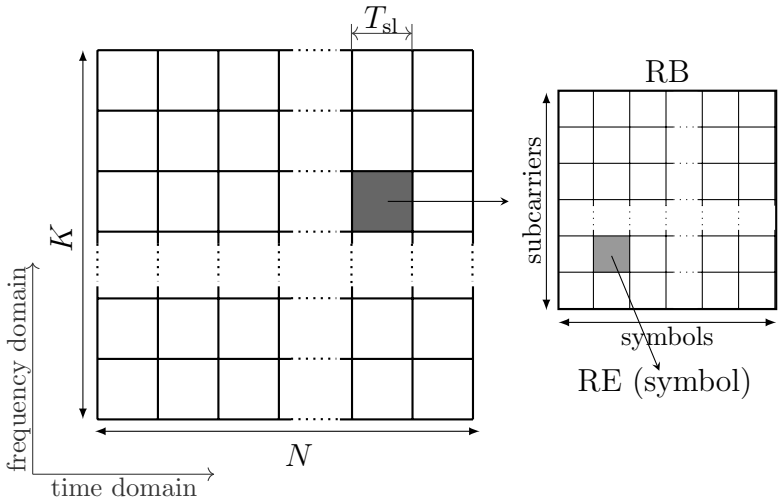


Figure 2.4: Time-frequency radio resource grid.

total number of symbols referred to as resource elements (REs) in each RB. Note that the subsequently considered system performance metrics depend exclusively on the number K_{RE}, not the specific form of the RB in terms of the number of symbols and subcarriers.

Moreover, the system is assumed to support A MCSs where the a^{th} MCS with $a \in \Omega_{1,A}$ is associated with a code rate of $R(a)$ and $S(a)$ coded bits per symbol. Furthermore, let $\beta(a, \gamma)$ be the resulting block-error rate (BLER) for the a^{th} MCS and for a given value γ of the signal-to-interference-plus-noise ratio (SINR)[1] over the considered V2V link where the SINR will be defined in Sect. 2.5. In this work, the BLER performance is assumed to be known.

For a sidelink transmission in a multicast group in a given slot[2], one or more RBs a MCS are assigned to the multicast group by the radio resource scheduler. Then, the transmitter V-UE of the multicast group performs sidelink multicast transmission to the other V-UEs in the group over a V2V link in the allocated RBs with the allocated MCS. Let $\mathbf{\Pi} = \{\Pi_{m,k}\}$ be an $(M \times K)$-dimensional resource allocation matrix with

$$\Pi_{m,k} = \begin{cases} a & \text{if RB } k \text{ and MCS } a \text{ are assiged to the } m^{\mathrm{th}} \text{ multicast group} \\ 0 & \text{otherwise.} \end{cases}$$

If RB k and MCS a are assigned to the m^{th} multicast group, i.e., if $\Pi_{m,k} = a$, the transmitter V-UE of the group $d_{\mathrm{tx},m}$ performs multicast transmission to other V-UEs in the group $\mathcal{D}_{\mathrm{rx},m}$ over the V2V links $\{\mathrm{L}_{d_{\mathrm{tx},m},\tilde{d}} \in \mathcal{L} : \tilde{d} \in \mathcal{D}_{\mathrm{rx},m}\}$ in the k^{th} RB using the a^{th} MCS. On the other hand, $\Pi_{m,k} = 0$ indicates that no resource is allocated to the m^{th} multicast group and therefore V-UE $d_{\mathrm{tx},m}$ does not perform any sidelink transmission in the k^{th} RB.

[1] All SINR values are based on the energy per bit.

[2] Subsequently, for better readability, the dependency of the RB index k and the MCS index a on the slot index n is not always denoted explicitly.

2.5 Sidelink Channel Quality

Fig. 2.5 illustrates an example of sidelink transmissions over V2V links in two distinct multicast groups in a common time-frequency resource. The transmission of

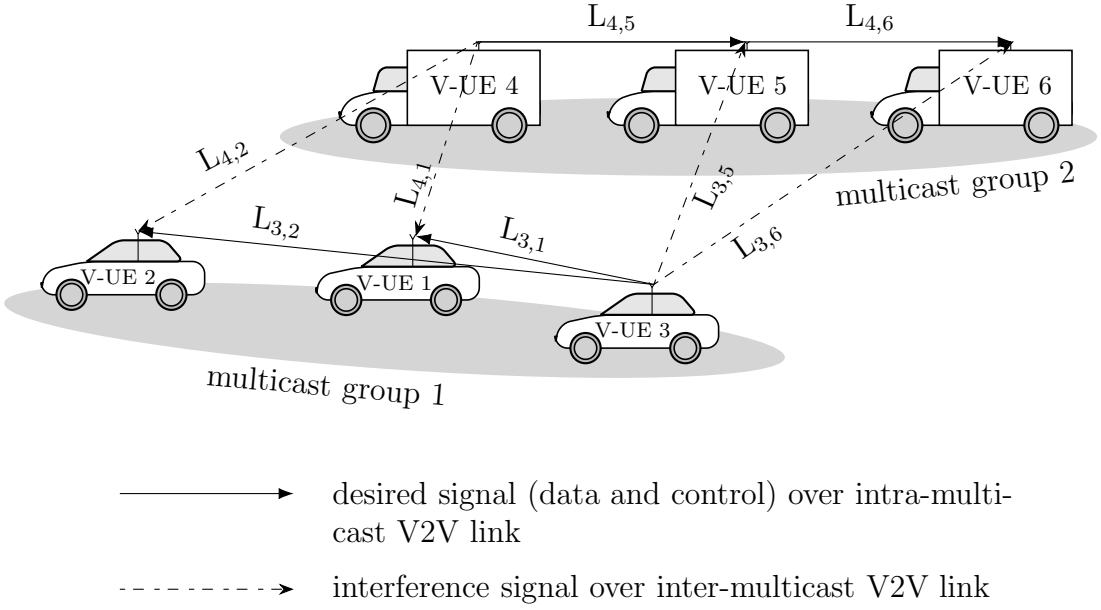


Figure 2.5: Illustration of desired V2V signal (data and control) transmission within multicast groups and interference signal across multicast groups.

the desired V2V signal (data and control) within each multicast group causes interference at the receiving V-UEs in the other multicast group.

The channel gain over the V2V link $\mathrm{L}_{d,\tilde{d}} \in \mathcal{L}$ in the k^{th} RB in time slot n is denoted as $H_{\mathrm{L}_{d,\tilde{d}},k,n} \in \mathbb{R}_0^+$ and contains the distance-dependent path loss $\eta_{\mathrm{L}_{d,\tilde{d}},n}$ and the normalized small-scale fading random variable $h_{\mathrm{L}_{d,\tilde{d}},k,n} \in \mathbb{R}_0^+$ where $H_{\mathrm{L}_{d,\tilde{d}},k,n} = \eta_{\mathrm{L}_{d,\tilde{d}},n} h_{\mathrm{L}_{d,\tilde{d}},k,n}$. Assuming flat fading in each RB for a transmission over link $\mathrm{L}_{d,\tilde{d}}$ with power $P_{\mathrm{L}_{d,\tilde{d}},k,n}$, the $\tilde{d}^{\text{th}}$ V-UE received signal power is $P_{\mathrm{L}_{d,\tilde{d}},k,n} H_{\mathrm{L}_{d,\tilde{d}},k,n}$. For a desired signal transmission over the V2V link $\mathrm{L}_{d,\tilde{d}}$ at time slot n in the k^{th} RB, the interference power received at the $\tilde{d}^{\text{th}}$ V-UE for a resource allocation $\mathbf{\Pi}$ reads

$$\sum_{\mathrm{L}_{d',\tilde{d}} \in \mathcal{L} \setminus \{\mathrm{L}_{d,\tilde{d}}\}} \mathbb{1}_{\Omega_{1,A}}\Big(\Pi_{\mu_{d'},k}\Big) P_{\mathrm{L}_{d',\tilde{d}},k,n} H_{\mathrm{L}_{d',\tilde{d}},k,n} \tag{2.2}$$

where $\mathbb{1}_{\Omega_{1,A}}\left(\Pi_{\mu_{d'},k}\right)$ is an indicator function with

$$\mathbb{1}_{\Omega_{1,A}}\left(\Pi_{\mu_{d'},k}\right) = \begin{cases} 1 & \text{if } \Pi_{\mu_{d'},k} \in \Omega_{1,A} \\ 0 & \text{if } \Pi_{\mu_{d'},k} \notin \Omega_{1,A}. \end{cases}$$

In this work, for simplicity, $P_{\mathrm{L}_{d,\tilde{d}},k,n}$ is set to P_{data} and P_{ctrl} for data and control signaling, respectively.

Moreover, for a zero-mean Gaussian thermal noise process with power spectral density σ^2, the SINR over the V2V link $\mathrm{L}_{d,\tilde{d}}$ at time slot n in the k^{th} RB is defined as

$$\gamma_{\mathrm{L}_{d,\tilde{d}},k,n}\left(\mathbf{\Pi}\right) = \frac{\mathbb{1}_{\Omega_{1,A}}\left(\Pi_{\mu_d,k}\right) P_{\mathrm{L}_{d,\tilde{d}},k,n} H_{\mathrm{L}_{d,\tilde{d}},k,n}}{\sigma^2 + \sum\limits_{\mathrm{L}_{d',\tilde{d}} \in \mathcal{L} \setminus \{\mathrm{L}_{d,\tilde{d}}\}} \mathbb{1}_{\Omega_{1,A}}\left(\Pi_{\mu_{d'},k}\right) P_{\mathrm{L}_{d',\tilde{d}},k,n} H_{\mathrm{L}_{d',\tilde{d}},k,n}}. \tag{2.3}$$

Clearly, the SINR over a V2V link is related to the channel quality of the link. Here, the sidelink performance of multicast transmission is limited by the link with the worst SINR in the multicast group. Therefore, the channel quality for the m^{th} multicast group transmission at time slot n defined by the V2V link of minimum channel quality in the group is given by

$$\gamma_{m,k,n}(\mathbf{\Pi}) = \min_{\tilde{d} \in \mathcal{D}_{\mathrm{rx},m}} \left\{ \gamma_{\mathrm{L}_{d_{\mathrm{tx},m},\tilde{d}},k,n}\left(\mathbf{\Pi}\right) : \mathrm{L}_{d_{\mathrm{tx},m},\tilde{d}} \in \mathcal{L} \right\}. \tag{2.4}$$

2.6 QoS Parameters

The sidelink resource allocation directly affects the QoS performance of sidelink communication. Here, reliability, transmission rate and packet delay are considered the key QoS parameters on V2V links for CAD sidelink communication thanks to their significance in CAD operation efficiency.

In this work, the reliability $\rho_{m,k,n}$ for V2V transmission within the m^{th} multicast group is viewed as the ability to receive the transmitted bits successfully in the k^{th} RB at time slot n within the group. Therefore, given the channel quality, the reliability can be expressed in terms of the BLER as

$$\rho_{m,k,n}(\mathbf{\Pi}) = 1 - \beta(\Pi_{m,k}, \gamma_{m,k,n}(\mathbf{\Pi})). \tag{2.5}$$

In addition, the minimum reliability achieved by V2V transmissions within the m^{th} multicast group at time slot n is defined by

$$\rho_{m,n}(\mathbf{\Pi}) = \begin{cases} \min\limits_{k\in\Omega_{1,K},\Pi_{m,k}>0} \rho_{m,k,n}(\mathbf{\Pi}) & \text{if } \sum\limits_{k\in\Omega_{1,K}} \Pi_{m,k} > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, an MCS-based rate calculation is adopted instead of Shannon capacity to reflect a practical system implementation with realistic calculation of the rate against the maximum theoretical limit (as represented by the Shannon limit). The transmission rate (or *rate* for short in the following) at time slot n for the m^{th} multicast group can be obtained as

$$r_{m,n}(\mathbf{\Pi}) = K_{\text{RE}} \sum_{k\in\Omega_{1,K}} S(\Pi_{m,k})\, R(\Pi_{m,k})$$

with a code rate $R(0) = 0$ and $S(0) = 0$ coded bits per symbol being defined for convenience. Then, with (2.1), the head-of-line (HOL) packet delay $\tau_{m,n}(\mathbf{\Pi})$ at the end of time slot n reads

$$\tau_{m,n}(\mathbf{\Pi}) = \begin{cases} \breve{\tau} & \text{if } r_{m,n}(\mathbf{\Pi}) < Q_{m,n} \\ 0 & \text{if } r_{m,n}(\mathbf{\Pi}) \geq Q_{m,n} \end{cases} \quad \text{with}$$

$$\breve{\tau} = \max\left\{ \tau \in \{1, ..., \Delta_m\} : r_{m,n}(\mathbf{\Pi}) - \sum_{\delta=\tau-1}^{\Delta_m-1} q_{m,\delta,n} < 0 \right\}.$$

Here, the HOL packet refers to the packet which is foremost in the queue, i.e., the packet which is waiting for the longest to be dequeued. Correspondingly, the HOL packet delay refers to the queuing delay associated with the HOL packet [53]. Since only the queueing delay component in the packet delay can be controlled by the radio resource scheduler, this work treats only the queueing delay and makes use of the queue's HOL packet delay, referred to as *delay* for short henceforth, for dynamic resource allocation.

2.7 QoS Model

In order to model QoS requirements of CAD applications with diverse elasticity levels of QoS parameters (i.e., rate, reliability and delay), utility functions are considered subsequently. In this regard, the level of satisfaction in the m^{th} multicast group using radio resources allocated to meet the QoS requirements of the multicast CAD V2V transmission is characterized by a utility function $U_m(r, \rho, \tau)$ depending on rate r, reliability ρ and delay τ. Here, $U_m(r, \rho, \tau)$ can be viewed as an aggregated utility (e.g., an appropriate weighted addition and/or multiplication [54]) of single criterion utility functions $u_{\text{rate},m}(r)$, $u_{\text{rel},m}(\rho)$ and/or $u_{\text{del},m}(\tau)$. Clearly, the rate and reliability utilities are monotonically increasing functions, while the delay utility is monotonically decreasing. Note that for a resource allocation $\mathbf{\Pi}$, the utility (i.e., level of satisfaction) at the m^{th} multicast group at time slot n is given by $U_m(r_{m,n}(\mathbf{\Pi}), \rho_{m,n}(\mathbf{\Pi}), \tau_{m,n}(\mathbf{\Pi}))$. QoS-based utilities for CAD will be discussed in Sect. 3.1.3.

Chapter 3

Cloud-Enabled Centralized Sidelink Radio Resource Management for CAD

CAD involves a group of vehicles controlled by a V2X application to exchange V2X messages for carrying out operations like CLC and platooning. The application is linked with vehicular dynamic models and control, and the V2X message generated by the application includes co-operative information for formulating and solving a CAD control problem for the considered use case [4]. When V2X messages are exchanged using direct V2V communications, the CAD performance (e.g., fuel efficiency in platooning) depends on the sidelink QoS performance (e.g., latency) as described in Chapter 1. Therefore, the required QoS performance on a sidelink for V2X message exchange is to be ensured to meet the targeted CAD performance.

In general, the sidelink RRM plays a key role in meeting the QoS requirements on V2V communication links. It involves management of radio spectrum, co-channel interference and transmission parameters such as MCS and transmission power for sidelink communication. By means of RRS, a V-UE can be assigned radio resources including radio spectrum resources (or RBs), transmission power and an MCS for sidelink transmission. This allows to control the data rate, the interference management as well as the link adaptation. Depending on the objective, e.g., to maximize

spectrum efficiency, to minimize latency on V2V links or to meet QoS requirements on key performance indicators (KPIs) etc., radio resource allocation strategies for RRS can be designed. As a result, upon definition of QoS requirements for CAD, suitable RRS strategies can be designed.

In a multi-operator environment, V-UEs may belong to different MNOs, where the sidelink transmission of a V-UE is controlled by the subscribed MNO. Consequently, V-UEs involved in a CAD operation such as CLC, sensor sharing and safety critical CCA may belong to different MNOs. Therefore, V2V interactions and interference coordination must be ensured among V-UEs across MNOs. In addition, sidelink RRM may need to be abstracted or possibly decoupled from the network of a specific MNO to allow sidelink radio resource control by vertical partners (e.g., car manufacturers) and applications. Furthermore, to perform QoS-aware sidelink RRS, the sidelink RRM entity of the network must be aware of the V2X application requirements and the involved V-UEs in CAD. In addition, it is essential for an application to adapt the CAD operation (e.g., increase IVD in platooning) by monitoring changes in the sidelink performance (e.g., increase in latency). Hence, there is a need for interaction between the sidelink RRM and the CAD application.

To this end, this chapter introduces a cloud-enabled sidelink RRM framework and proposes centralized sidelink RRS schemes to support CAD. In the following, Sect. 3.1 describes the framework that allows for placement of sidelink RRM in a logically centralized cloud entity for sidelink RRM abstraction from a network. The approach allows the sidelink RRM to interact with a V2X application, which includes capabilities of the V2X application to indicate QoS requirements. Subsequently, Sect. 3.2 discusses QoS-aware sidelink RRS schemes within the proposed framework, where a QoS requirement-aware sidelink resource allocation optimization problem for CAD is formulated and analyzed for cloud-based sidelink RRM in a dynamic vehicular environment.

3.1 Radio Resource Management Framework

For CAD operation, vehicles can use wireless communication services, which are provided by the MNOs to perform V2V communication, to exchange CAD information over V2V links. Each MNO owns or controls the necessary elements such as wireless network infrastructure (e.g., BSs) and radio spectrum that are needed to enable wireless communication. V-UEs belonging to different operators are typically assumed to transmit using different spectral bands (i.e., different carrier frequencies). However, this precludes the possibility of spectrum sharing among operators.

V2V interactions and interference coordination must be ensured among V-UEs across MNOs to support CAD. Hence, future spectrum allocation for V2X sidelink communication may not belong to a specific operator, but the sidelink radio resources may need to be shared among MNOs. However, currently, each MNO has its own sidelink radio resource scheduler located at its BSs which make use of measurements such as CSI and geographical location to ensure that orthogonal time-frequency resources are assigned to nearby V-UEs to prevent interference within the respective coverage area. Clearly, these parallel MNO-specific approaches are unaware of V-UEs belonging to other MNOs and measurements associated with these V-UEs. Consequently, if nearby V-UEs belonging to different operators transmit on the same carrier frequency, the operator-centric sidelink schedulers are unable to prevent interference. In addition, the operator-centric sidelink schedulers cannot avoid packet losses resulting from asynchronous half-duplex transmissions. Thus, there is a need for managing sidelink radio resources in an operator-independent manner. Furthermore, the sidelink RRM may need to be abstracted or possibly decoupled from the operator's network to allow sidelink radio resource control by vertical partners such as automotive organizations and applications. Here, the vertical partners may be in need of controling sidelink communication, in particular, if

they have a dedicated radio spectrum for their applications. Hence, vertical partners may require the sidelink RRM to be decoupled from the operator's network. Moreover, the need for QoS-aware sidelink RRM necessitates interaction between the sidelink RRM and the CAD application. To this end, a cloud-enabled RRM framework is proposed in the following.

3.1.1 Cloud-Based Sidelink Radio Resource Scheduler

The cloud-enabled RRM framework proposes to move the intelligence (control of sidelink radio resources, including sidelink RRS) from the operator's RAN to a network entity, a cloud server, that is accessible by different MNOs. The cloud server comprises software or a processor that is configured for sidelink RRM including RRS on behalf of one or more MNOs. This allows for RRM across networks of different MNOs in an operator-independent manner. Furthermore, since an external access to the sidelink RRM in the cloud server can be enabled by means of APIs, the framework also allows for radio resource control by external applications and vertical partners.

In this work, a cloud server is considered to be a logically centralized network entity consisting of one or more remotely placed physical servers and being accessible by the V-UEs via BSs over V2N links. Since the sidelink RRM functionality is placed at the cloud server, the sidelink RRS is performed in the proposed framework at the central cloud server. The V2N links carry control plane information related to the sidelink RRM such as, e.g., sidelink measurements and schedule grants. The sidelink RRM function configures V-UEs to report sidelink channel measurements (e.g., CSI) and packet buffer status, i.e., queue state information (QSI), periodically to the cloud server. Here, the QSI report from a V-UE, which implicitly indicates a non-empty buffer (or queue), is seen as a *schedule request* for sidelink radio resources from the V-UE to the radio resource scheduler. Upon RRS, the scheduler provides a

schedule grant to V-UEs indicating the radio resources (e.g., RBs and/or MCSs) for sidelink transmissions. Fig. 3.1 illustrates the cloud-based RRS. Here, the interaction

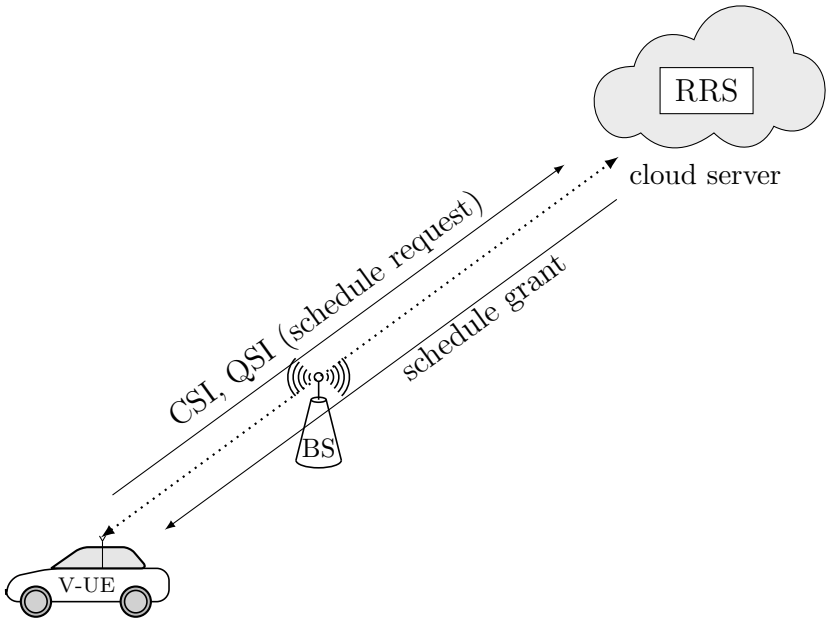


Figure 3.1: Illustration of cloud-based RRS.

between the RRM function in the cloud server and the V-UEs can be implemented in either of two ways.

In one implementation, a new communication interface can be introduced between the cloud server and the BSs belonging to different MNOs [44] operating, e.g., 5G mobile radio systems. Then, the cloud server communicates with an operator's BSs via the new interface to control the radio resources for sidelink communication within the coverage area of all BSs. The BSs and V-UEs exchange control information related to sidelink communication as usual (or unmodified) via the already existing standard interface (RRC $\leftrightarrow$ medium access control (MAC) $\leftrightarrow$ physical layer) specified in, e.g., 5G. Thus, the V-UE is unaware that sidelink radio resources are controlled from outside the RAN. A key advantage of this implementation is that no changes to the underlying network interfaces between V-UE and the BS or within the V-UE architecture are required.

In an alternative implementation, control plane signaling related to the control of radio resources of a sidelink communication is carried over a user plane. This can be done, in particular, through Internet protocol (IP) encapsulation of the control plane over the underlying communication system (e.g., 5G) between the cloud server and the V-UEs [45]. Here, the sidelink RRM entity and V-UEs are considered to comprise an IP unit that is configured to encapsulate RRC messages within an IP packet and are able to exchange the IP packet over IP on V2N links.

In the following, an implementation of cloud-based sidelink RRM is discussed where the sidelink RRM functionality consisting of the radio resource scheduler is placed in a V2X application server.

3.1.2 V2X Application Layer

Fig. 3.2 illustrates the application layer interaction of a V-UE (V-UE 1) with a V2X application server and another V-UE (V-UE 2). Here, the V2X application server is

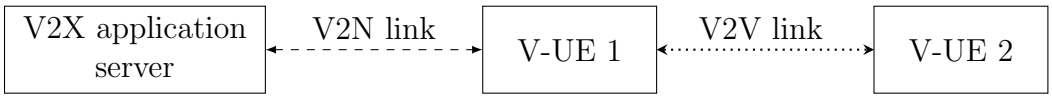


Figure 3.2: Overview of V2X application layer interaction.

a cloud server and hosts one or more V2X application-specific server(s) (e.g., platoon servers) configuring and/or handling application(s) (e.g., platooning) in V-UEs. The V2X applications are associated with vehicular dynamic models and control. Furthermore, they carry out CAD operations like, e.g., CLC and platooning, i.e., they formulate and solve CAD control problems as per the use cases. To perform CAD, these applications generate or process V2X messages with co-operative information that is required to formulate (based on, e.g., vehicle state, sensor data, trajectories) and to solve (using, e.g., messages in a distributed iterative algorithm) the control problem. Furthermore, they are also responsible for choosing appropriate

LoA. V-UEs communicate with V2X application servers using V2N links provided by the underlying communication system (e.g., 5G) and use V2X application-specific client(s) to exchange application-level signals with V2X application-specific server(s) over V2N links. Moreover, V-UEs use V2V links to exchange V2X messages with other V-UEs.

As noted above, to support CAD, there is a need for interaction between the RRM function and the V2X applications for QoS-aware RRS as well as to allow for V2X application adaptation based on KPIs such as latency and reliability. Hence, this work places the sidelink RRM functionality consisting of radio resource scheduler in V2X application server as shown in Fig. 3.3 and provides an interface through

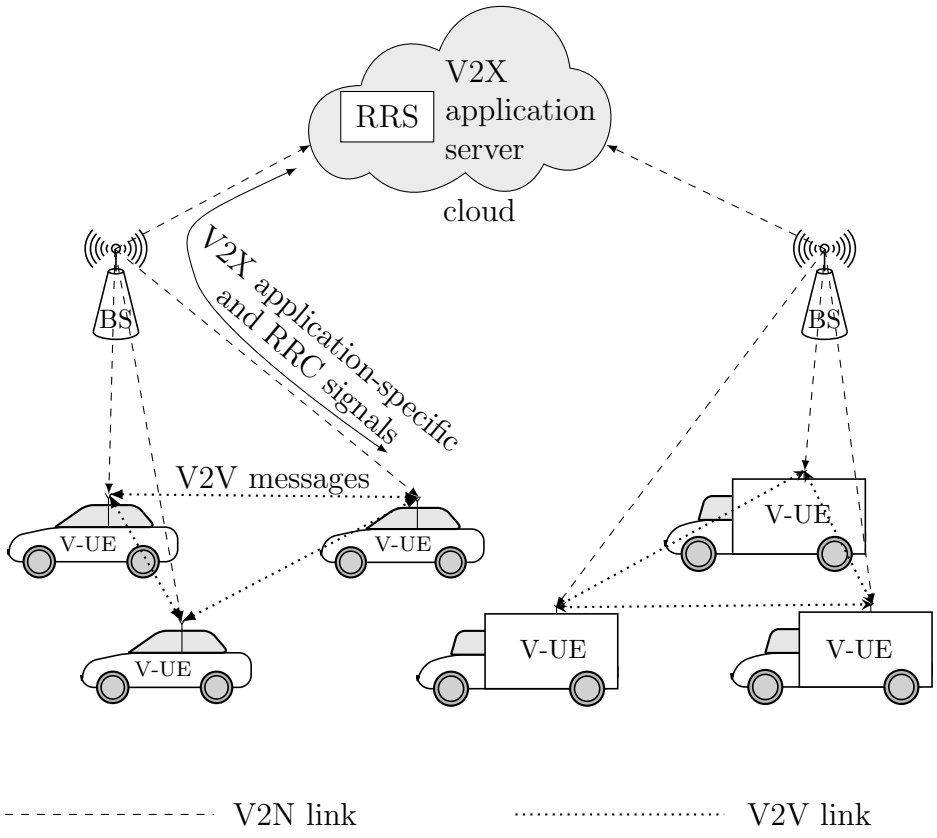


Figure 3.3: Illustration of sidelink RRS entity placement in a cloud-based V2X system.

APIs between sidelink RRM function and the V2X application-specific server(s)

(e.g., platoon server). With this interface, the sidelink RRM entity can receive and maintain information on QoS requirements (as described in Sect. 3.1.3) and vehicular groups (CAD groups and multicast groups as described in Sect. 2.2) associated with the CAD operation from the V2X application-specific server(s).

In 3GPP specified 5G networks, the application layer functional support to V2X services offered by the 3GPP can be exploited to realize the aforementioned interface between V2X application-specific server and the sidelink RRM function. In the functional support, 3GPP offers service enabler architecture layer (SEAL) services such as location management, group management (GM) (e.g., group creation and deletion) and network resource management (NRM) (e.g., unicast, multicast resources) via V2X application enabler (VAE) to the V2X applications in 5G [55], [56]. These services are provided through APIs by SEAL servers which reside in the V2X application server and can communicate with the underlying 3GPP network systems using the respective 3GPP interfaces. Then, by placing the sidelink RRM entity in the SEAL servers, APIs existing between the SEAL servers and the V2X application-specific server can be utilized as an interface between the sidelink RRM entity and the V2X application-specific server. More details on the 3GPP application layer functional support to V2X services are described in Appendix B.

In the following, the scope of the work is limited to sidelink RRS and it is assumed that the V2X application-specific server dynamically adapts CAD V2X applications by monitoring the network performance while indicating the desired QoS requirements to the sidelink RRM entity periodically through the given API.

3.1.3 CAD QoS Requirements

CAD QoS requirements span a wide range with different levels of elasticity depending on the use case, level of criticality and LoA. Here, the elasticity of the QoS requirement reflects the QoS performance tolerance of the CAD application,

however, at the cost of suboptimal CAD operations w.r.t., e.g., fuel and/or road efficiency. For example, a large delay in V2V communications may lead to a larger IVD due to an increased IVD safety gap which results in lower fuel efficiency gain in platooning [6]. Hence, at a given LoA, a platooning application may be able to tolerate a higher latency at the cost of a reduced fuel efficiency gain. It is noted that beyond a certain latency the efficiency gain may be insignificant and therefore the application may choose to terminate the platooning or to lower the LoA. Here, the ability to tolerate a QoS performance degradation also depends on the LoA. At a higher LoA, the platooning application is more stringent with regard to QoS tolerance as the reliance on ADS increases for increasing LoA such that the degradation in QoS performance increases the risk of a collision. Furthermore, an increased QoS performance may not translate to a proportionate increase in fuel efficiency gain for platooning. Hence, for example, a change in fuel efficiency is greater for a change in latency performance of a high-latency V2V link than that of a low-latency V2V link. The criticality of the CAD operation also dictates the QoS performance. For example, in case of CCA, the application may have very low tolerance to QoS degradation as the reduced QoS performance significantly increases the collision (or accident) risk for vehicles. Hence, the CCA application may have a strict minimum QoS requirement. Considering that the sidelink RRS directly impacts the QoS performance, it is crucial that the CAD QoS requirements are adequately captured and conveyed to the sidelink RRM entity to allow for QoS-aware RRS.

Utility functions are known to be an effective means to capture QoS requirements where the utility[1], as a metric, reflects the user satisfaction for a given network performance [57]–[59]. Few notable utility functions are logarithmic, sigmoid, piecewise step and linear functions as illustrated in Fig. 3.4. By carefully choosing the utility function, different QoS requirements with diverse elasticity level can be captured.

[1]The notions *utility* and *utility function* are used interchangeably in this work.

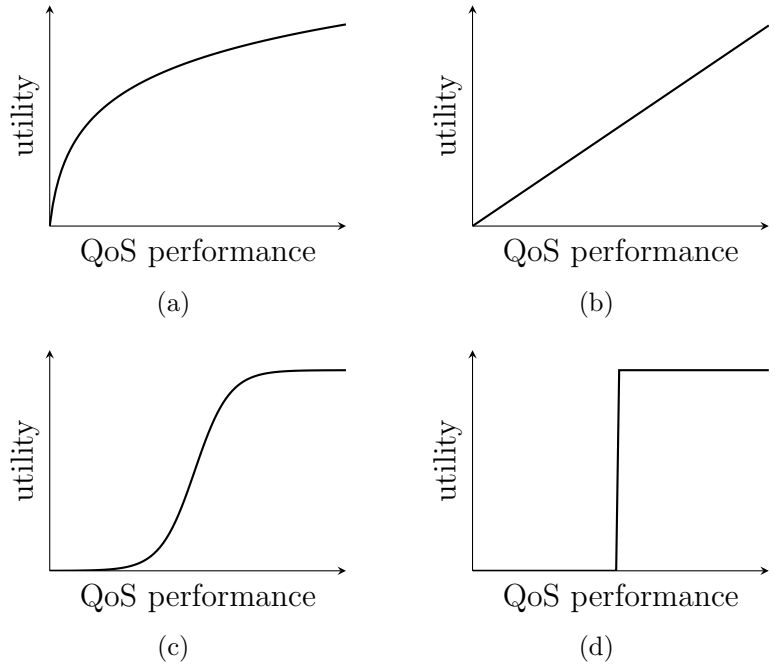


Figure 3.4: Illustration of example utility functions: (a) logarithmic, (b) linear, (c) sigmoid and step utility functions.

Therefore, considering the diverse elasticity levels of QoS (rate, reliability and delay) requirements of CAD, this work proposes to capture the CAD application requirements using utility functions. That is, the level of satisfaction in a CAD group, hence in a multicast group, for the QoS performance (rate r, reliability ρ and delay τ) on a V2V link is characterized by the utility function $U(r, \rho, \tau)$. Here, $U(r, \rho, \tau)$ can be viewed as an aggregated utility (e.g., an appropriate weighted addition, multiplication [54]) of single-criterion utility functions $u_{\text{rate}}(r)$, $u_{\text{rel}}(\rho)$ and/or $u_{\text{del}}(\tau)$. Clearly, the rate and reliability utilities are monotonically increasing functions, while the delay utility is monotonically decreasing.

Utility functions for multicast groups are considered to be provided by the V2X application-specific server based on the elasticity level of the QoS requirements of the corresponding CAD application and the LoA. Furthermore, since the sigmoid utility function effectively models soft real-time applications by reflecting user sensitivity to the changes in the criterion value through a steepness parameter [54], a

single-criterion utility can be modeled as a sigmoid function to capture diverse elasticity level requirements of CAD. Here, as the steepness parameter increases, the utility curve becomes steeper representing a higher sensitivity of the application. As an alternative, the utility can be modeled as a step function in case of a restriction on the criterion value and as a logarithmic function for the case of elastic traffic. Furthermore, revisiting network criteria being considered in the network utility maximization (NUM) framework [57] and its extensions, the following aspects are to be taken into account to find a suitable utility function $U(\cdot)$ for service requirements of multicast transmissions supporting CAD.

- $U(r)$: A utility as function of the rate r is introduced in the basic NUM framework where the links are assumed to provide fixed rates. Since a large amount of data needs to be shared among vehicles under the use cases of co-operative sensing (sensor sharing, see-through applications), rate-based utilities can provide a good QoS measure considering bounded reliability and delay. Furthermore, the possibility to operate under variable video resolutions and formats makes utility functions such as logarithm and sigmoid more suitable, where the choice of the utility function depends on the elasticity requirements. However, the rate r alone does not sufficiently capture the user needs for use cases, in particular, under co-operative maneuvers. To this end, extensions of the basic NUM considering delay and/or reliability are to be used.

- $U(r, \rho)$: A utility as a function of rate r and reliability ρ allows for flexible utilization of rate-reliability characteristics at the physical layer to support trade-offs between rate and reliability [58]. In case of co-operative awareness and co-operative maneuvers in emergency situations (e.g., emergency lane change, emergency brake in dense platooning, forward collision warning), a higher reliability is required and the data packet size can be flexibly modified to include

emergency notification messages and/or planned trajectories. Thus, for such use cases depending on the elasticity or strictness of the reliability requirements, a sigmoid function or a piecewise step function can be used along with the aforementioned rate utility to capture rate-reliability trade-offs. However, even though this can be constrained by a delay bound, it falls short of capturing delay requirements.

- $U(r, \rho, \tau)$: The delay τ can be incorporated along with the rate-reliability into the utility function [59]. Co-operative maneuvers, in particular, can be operated under lenient delay requirements at the cost of a sub-optimal performance (e.g., fuel efficiency in dense platooning, road efficiency in CLC). The lower the latency, the denser can be the platoons with a smaller IVD, the higher will be the fuel efficiency, the quicker is the lane change and the larger is the road efficiency. Based on the criticality of the service, a sigmoid can help to model flexible delay requirements in co-operative driving due to its ability to capture delay sensitivity. Therefore, considering the delay τ with rate-reliability has the potential to describe flexible QoS requirements of CAD.

3.2 Full CSI-based Dynamic Resource Scheduling

To exploit the cloud-enabled sidelink RRM framework for RRS to support CAD V2V communication in a multi-operator environment, a centralized RRS scheme is presented where the centralized scheduler allocates radio resources to V-UEs in order to maximize the aggregated utility across multicast groups by acquiring full CSI (or instantaneous CSI) and QSI from the V-UEs.

3.2.1 Problem Description

CAD is translated to a combination of wide range of QoS (i.e., rate, reliability and/or delay) requirements on multicast group communication in CAD groups depending on the nature of CAD and the operating LoA as elaborated in Sect. 1.1.1. Therefore, the sidelink RRS must take into account the diverse elasticity levels of QoS requirements of CAD. With the QoS requirements of the applications being modeled in terms of the chosen utility function as described in Sect. 3.1.3, the aim of the RRS is to maximize the utility. In this regard, the NUM framework [57] and its extensions have been widely used in resource allocation [59].

The resource allocation for CAD V2V communications is viewed as an optimization problem to maximize the aggregated QoS-dependent utility [46], i.e., the sum of individual utilities of multicast groups, thereby finding a rate-reliability-delay trade-off across multicast groups. Then, with $\mathcal{F}_n = \{m : Q_{m,n} > 0\}$ being the set of all multicast groups with non-empty queues at time slot n, the resource allocation problem can be formulated as

$$\hat{\mathbf{\Pi}}_{\text{opt},n} = \arg \max_{\mathbf{\Pi} \in \Omega_{0,A+1}^{M \times K}} \sum_{m \in \mathcal{F}_n} U_m(r_{m,n}(\mathbf{\Pi}), \rho_{m,n}(\mathbf{\Pi}), \tau_{m,n}(\mathbf{\Pi})) \tag{3.1}$$

subject to

$$\Pi_{m,k} = 0 \quad \forall\, (m,k) \in \Omega_{1,M} \setminus \mathcal{F}_n \times \Omega_{1,K} \tag{3.1a}$$

$$\sum_{m \in \mathcal{M}_{\text{CAD},x}} \max_{k \in \Omega_{1,K}} \mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) \leq 1 \quad \forall x \in \Omega_{1,X}. \tag{3.1b}$$

Here, (3.1a) assures that the resources are exclusively allocated to multicast groups with non-empty queues to avoid resource wastage. Since this work considers each multicast transmission from a member V-UE to be directed to all other member V-UEs within the respective CAD group, the transmitter V-UE will not be able to

receive any multicast transmissions performed concurrently with its own transmission within the group due to a half-duplex constraint at the V-UE. Hence, considering the half-duplex constraint, (3.1b) avoids resource allocation to more than one multicast group of the same CAD group at any time slot.

This work focuses on the delay and reliability performance of CAD V2V communications, where the sidelink resource allocation takes into account the delay and reliability requirements of the CAD groups. Here, an inelastic (strict) requirement on reliability is assumed where $\rho_{\text{req},m}$ indicates the reliability requirement on the V2V transmission for the m^{th} multicast group. Moreover, an elastic requirement on the delay is assumed and is captured by a delay-dependent single-criterion utility function $u_{\text{del},m}(\cdot)$. The V2X application-specific server transmits $\rho_{\text{req},m}$ and $u_{\text{del},m}(\cdot)$ to the sidelink RRM entity periodically and is assumed to adapt the packet generation rate at the CAD applications by monitoring the sidelink performance.

A resource allocation $\mathbf{\Pi}$ must ensure that $\rho_{m,n}(\mathbf{\Pi}) \geq \rho_{\text{req},m}$ to meet the reliability requirements of the m^{th} multicast group. That is, the minimum reliability achieved by V2V transmissions within the m^{th} multicast group in the k^{th} RB at time slot n needs to be at least $\rho_{\text{req},m}$, i.e.,

$$\rho_{m,k,n}(\mathbf{\Pi}) \geq \rho_{\text{req},m}. \tag{3.2}$$

It is noted from (2.5) that the reliability performance of a V2V transmission relies on the SINR performance on the corresponding V2V link. Therefore, the reliability requirement can be translated into an SINR requirement where a multicast V2V transmission meets the reliability requirement $\rho_{\text{req},m}$ if the channel quality for the m^{th} multicast group is at least $\gamma_{\text{thr},m}(a)$ for the a^{th} MCS. Hence, the reliability

constraint in (3.2) can be written in terms of the channel quality as

$$\gamma_{m,k,n}(\mathbf{\Pi}) \geq \gamma_{\mathrm{thr},m}(\Pi_{m,k}). \tag{3.3}$$

Then, considering the delay requirement in terms of the utility function $u_{\mathrm{del},m}(\cdot)$ and the reliability requirement in terms of the channel quality requirement (3.3) in (3.1), the resource allocation problem in (3.1) can be reformulated as

$$\hat{\mathbf{\Pi}}_{\mathrm{opt},n} = \arg \max_{\mathbf{\Pi} \in \Omega_{0,A+1}^{M \times K}} \sum_{m \in \mathcal{F}_n} u_{\mathrm{del},m}\left(\tau_{m,n}(\mathbf{\Pi})\right) \tag{3.4}$$

subject to

$$\Pi_{m,k} = 0 \quad \forall\,(m,k) \in \Omega_{1,M} \backslash\, \mathcal{F}_n \times \Omega_{1,K} \tag{3.4a}$$

$$\sum_{m \in \mathcal{M}_{\mathrm{CAD},x}} \max_{k \in \Omega_{1,K}} \mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) \leq 1 \quad \forall x \in \Omega_{1,X} \tag{3.4b}$$

$$\gamma_{\mathrm{thr},m}(\Pi_{m,k}) - \gamma_{m,k,n}(\mathbf{\Pi}) \leq 0 \quad \forall (m,k) \in \mathcal{F}_n \times \Omega_{1,K}. \tag{3.4c}$$

While the constraints (3.4a) and (3.4b) are identical to (3.1a) and (3.1b), resp., (3.4c) assures that the reliability requirement of multicast groups are satisfied by meeting (3.3).

3.2.2 Solution

Since (3.4) is a constrained combinatorial optimization problem, this work proposes a solution embodying iterative algorithms with low complexity being described subsequently. At first, **Algorithm** 1 attempts to reduce the solution search space by nominating at most one multicast group for resource allocation from each CAD group. Then, **Algorithm** 2 iteratively assigns RBs and MCSs to nominated multicast groups considering the corresponding delay and reliability requirements.

As per the constraint (3.4b), at most one multicast group can be scheduled within a CAD group at any time slot to avoid packet loss at one or more member V-UEs due to the half-duplex constraint. Therefore, instead of considering all multicast groups $\mathcal{F}_n$ with non-empty queues for resource allocation, optionally the RRS can consider only a subset of $\mathcal{F}_n$ in (3.4) by assigning the most preferable multicast groups in $\mathcal{M}_{\text{CAD},x}$ for scheduling in the x^{th} CAD group at time slot n. This significantly reduces the search space since at most only a predefined number X of multicast groups are considered for resource allocation instead of M multicast groups. Let $\check{\mathcal{F}}_n \subseteq \mathcal{F}_n$ be the set of all selected multicast groups with non-empty queues at time slot n. Here, since the packet delay dependent utility function $u_{\text{del},m}(\cdot)$ is identical for all $m \in \mathcal{M}_{\text{CAD},x}$, the x^{th} CAD group can nominate a multicast group that would return the maximum utility at the given time slot (i.e., the multicast group whose HOL delay at the end of the time slot corresponds to a steeper slope on the utility function than that of all other multicast groups in the CAD group) as captured in **Algorithm** 1. Here, the algorithm considers the negative slope as the packet

Algorithm 1: Multicast Group Nomination in CAD Groups

Input : QSI with $\mathcal{F}_n$
Output: $\check{\mathcal{F}}_n$

Initialization: $\check{\mathcal{F}}_n = \emptyset$, $\mathbf{\Pi} \leftarrow \mathbf{0}_{M \times K}$;
for $x = 1 : X$ **do**
  $\mathcal{M} \leftarrow \mathcal{M}_{\text{CAD},x} \cap \mathcal{F}_n$;
  if $\mathcal{M} \neq \emptyset$ **then**
    $\acute{m} \leftarrow \arg\max_{\tilde{m} \in \mathcal{M}} -u'_{\text{del},\tilde{m}}\left(\tau_{\tilde{m},n}(\mathbf{\Pi})\right)$;
    $\check{\mathcal{F}}_n \leftarrow \check{\mathcal{F}}_n \cup \{\acute{m}\}$
  end if
end for

delay is a downward criterion and, for the given allocation $\mathbf{\Pi}$, the slope of the utility function for the HOL delay at the end of the current time slot τ is given by

$u'_{\mathrm{del},m}\left(\tau_{m,n}(\mathbf{\Pi})\right) = \frac{\mathrm{d}u_{\mathrm{del},m}(\tau)}{\mathrm{d}\tau}\Big|_{\tau=\tau_{m,n}(\mathbf{\Pi})}$.

For the nominated multicast groups in **Algorithm** 1, the resource allocation is performed according to **Algorithm** 2 assigning RBs and MCSs. Iterating over RBs,

Algorithm 2: RRS Algorithm

Input : full CSI and QSI with $\check{\mathcal{F}}_n$
Output: $\mathbf{\Pi}$

Initialization: $\mathbf{\Pi} \leftarrow \mathbf{0}_{M\times K}$;
for $k = 1 : K$ **do**
  $\mathcal{M} \leftarrow \check{\mathcal{F}}_n$;
  while $|\mathcal{M}| > 0$ **do**
    $\acute{m} \leftarrow \arg\max_{\tilde{m}\in\mathcal{M}} -u'_{\mathrm{del},\tilde{m}}\left(\tau_{\tilde{m},n}(\mathbf{\Pi})\right)$;
    for $a = A : 1$ **do**
      $\tilde{\mathbf{\Pi}} \leftarrow \mathbf{\Pi}$;
      $\tilde{\Pi}_{\tilde{m},k} \leftarrow a$;
      if $\gamma_{\mathrm{thr},\acute{m}}(\tilde{\Pi}_{\tilde{m},k}) - \gamma_{\acute{m},k,n}(\tilde{\mathbf{\Pi}}) \leq 0$ **then**
        $\Pi_{\acute{m},k} \leftarrow \tilde{\Pi}_{\tilde{m},k}$;
        break;
      end if
    end for
    $\tilde{\mathcal{M}} \leftarrow \mathcal{M} \setminus \{\acute{m}\}$;
    foreach $\tilde{m} \in \mathcal{M}$ **do**
      $\tilde{\mathbf{\Pi}} \leftarrow \mathbf{\Pi}$;
      $\tilde{\Pi}_{\tilde{m},k} \leftarrow 1$;
      if not $\gamma_{m,k,n}(\tilde{\mathbf{\Pi}}) > (2\mathbb{1}_{\Omega_{1,A}}(\tilde{\Pi}_{m,k}) - 1)\gamma_{\mathrm{thr},m}(\tilde{\Pi}_{m,k}), \forall m$ **then**
        $\tilde{\mathcal{M}} \leftarrow \tilde{\mathcal{M}} \setminus \{\tilde{m}\}$;
      end if
    end foreach
    $\mathcal{M} \leftarrow \tilde{\mathcal{M}}$;
  end while
end for

the algorithm assigns RBs incrementally to the multicast groups in $\check{\mathcal{F}}_n$ by selecting a

multicast group for allocation in each RB based on a utility-based greedy criterion. Here, the greedy criterion refers to the allocation strategy where the resources are allocated in each iteration in a locally optimal manner. In other words, in each iteration the algorithm allocates resources to maximize the resulting utility in the iteration. For the chosen multicast group, the algorithm also allocates the best MCS (i.e., with highest order MCS) that can be supported in the RB considering SINR performance and reliability requirements. Furthermore, upon assigning a multicast group with an MCS in the RB, the algorithm also attempts to iteratively accommodate as many multicast groups as possible in the RB with the best possible MCS as long as the SINR performance is within the acceptable limit (i.e., reliability requirements are met) for all the multicast groups assigned in the RB. Since the objective of (3.4) is to maximize the aggregated utility of all multicast groups in $\check{\mathcal{F}}_n$, the algorithm considers the negative slope (as above, the packet delay is a downward criterion) of the utility function $-u'_{\mathrm{del},m}\left(\tau_{m,n}(\mathbf{\Pi})\right)$ for the HOL delay at the end of the current time slot $\tau_{m,n}(\mathbf{\Pi})$ for the given allocation $\mathbf{\Pi}$ as a greedy criterion.

3.2.3 Advantages and Challenges

The radio resource allocation problem (3.4) is to maximize the aggregated utility across multicast groups. Therefore, finding the solution to the problem requires a centralized dynamic scheduler. Furthermore, the scheduler requires full CSI for appropriate MCS selection and QSI for delay calculation and identifying non-empty queues. Hence, each V-UE must provide to the dynamic scheduling entity a context information which includes instantaneous CSI measurements for the associated V2V links and QSI for the transmitting V-UE. Moreover, such a centralized solution requires placement of sidelink dynamic RRM functionalities at the central cloud server.

The proposed solution considers QoS requirements of multicast V2V communications in CAD groups by means of utility functions and aims to maximize the aggregated utility of all multicast groups. This brings a flexibility at the sidelink RRM function to dynamically adapt the RRS to the diverse resource demands from CAD operations and cater to them to maximize CAD operation efficiency. Here, the proposed solution benefits from the availability of QSI and full CSI. With QSI, the solution is aware of packet queuing delay and performs prioritization of multicast transmissions considering the respective multicast group delay requirement. Also, it avoids resource wastage by not allocating resources to multicast groups with non-empty queues. In addition, utilizing the acquired full CSI measurements, the solution takes into account the SINR performance of each multicast group while allocating resources so that the reliability performance is within the acceptable limit. The solution increases the resource reuse for higher resource utilization by accommodating as many multicast transmissions as possible in each RB while meeting the respective reliability requirement. It takes into account the half-duplex constraint at V-UEs and avoids resource allocation to more than one multicast groups of the same CAD group at any time slot. Furthermore, by performing the RRS at the central cloud server, the proposed solution inherits the advantages of cloud-enabled centralized sidelink RRM framework. In particular, the sidelink RRM can perform RRS to all V-UEs across MNOs and can be decoupled from MNOs' networks so that, e.g., vertical partners like automotive consortia can directly control the RRS for sidelink communication.

Dynamic scheduling performed at the central cloud server results in a high signaling overhead over the V2N link due to acquisition of full CSI and QSI from V-UEs. It is noted that the V2N links (or a part of the network infrastructure associated with V2N links) between the V-UEs and the cloud server may be managed and offered as a service by MNOs for sidelink RRM. For instance, an automotive orga-

nization or consortium that has a license to use a specific sidelink radio spectrum band may perform sidelink RRM within the band and may utilize V2N links offered by MNOs to exchange RRC information between the V-UEs and the cloud server in a *pay-per-use* manner. The resulting high signaling overhead over the V2N links increases operating costs for sidelink RRM and is therefore not desired.

When the V2N link traverses through a BS, it comprises the V2I link between the V-UE and the BS as well as the backhaul link between the BS and the cloud server. Here, the V2I link introduces a V2I link delay (or V2I delay for short) due to the propagation delay between the V-UE and the BS. In addition, any processing of V2N packets at the BS for the purpose of, e.g., forwarding the packets to the cloud server via the backhaul link incurs delay. Furthermore, the backhaul link introduces additional delay over the V2N link. Depending on the deployed backhaul topologies and technologies, the backhaul delay could be significantly larger due to a high transport network delay. As a result, the V2N delay between the V-UEs and the cloud server could be significantly large.

In the centralized dynamic RRS, the V-UE sends a schedule request over the V2N link. Subsequently, the schedule grant is provided to the V-UE by the cloud server over the V2N link up on performing the RRS. Therefore, a large V2N delay results in a higher scheduling delay as illustrated in Fig. 3.5 and consequently in a larger queuing delay. This could potentially lead to a sub-optimal CAD operation. Furthermore, the V-UE reports the CSI to the scheduler at the cloud server over the V2N link. But, as the vehicular radio environment is highly dynamic due to high mobilities of the involved vehicles, the large V2N delay may render the CSI available at the cloud server to be invalid. Fig. 3.6 illustrates the delay between CSI reporting and the schedule grant availability at a V-UE in a cloud-enabled centralized dynamic RRS. Here, the sidelink RRM function placed at the cloud server acquires full CSI. The latter is provided in form of a report containing instantaneous CSI measured

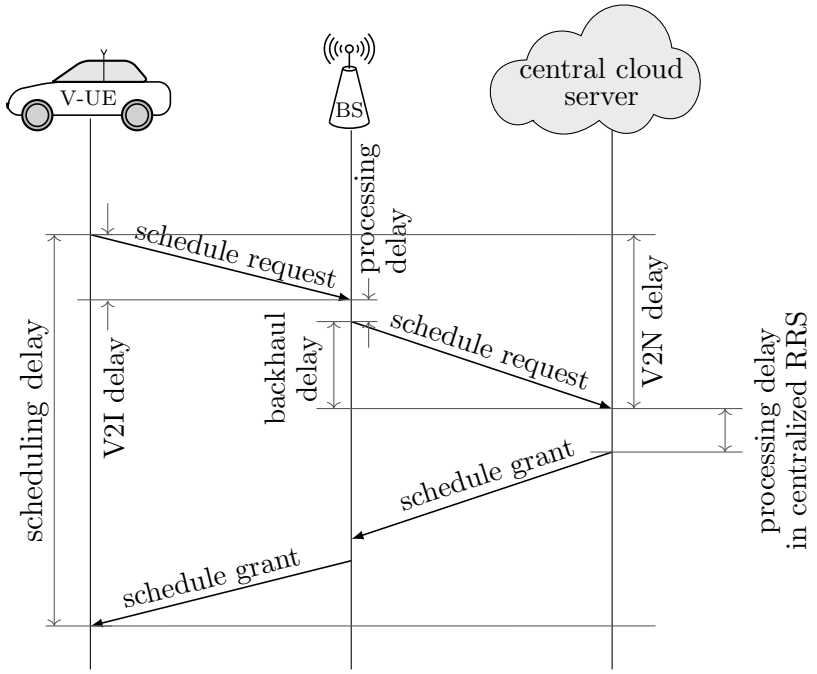


Figure 3.5: Illustration of scheduling delay in centralized dynamic scheduling at the cloud server.

at the V-UE for all the associated V2V links. In case of a large delay over the V2N link, by the time the RRS is performed at the cloud server, the vehicular channel conditions may have already changed. Also, the channel conditions may have significantly changed by the time the V-UE receives the schedule grant from the time it had reported CSI as well as the time the RRS was performed. Consequently, the usage of radio resources that are scheduled with invalid (expired) CSI may have significant adverse impact on the SINR, and therefore the reliability and performance on V2V links.

In case of PDT V2V transmissions, the need for explicit QSI reporting from all V-UEs at every time slot can be eliminated to reduce signaling overhead over V2N links as the resource demand for data transmission can be predicted with accuracy[2]. In this regard, a virtual queue can be implemented and maintained in the RRM to

[2]Clearly, this also applies to highly predictable data traffic

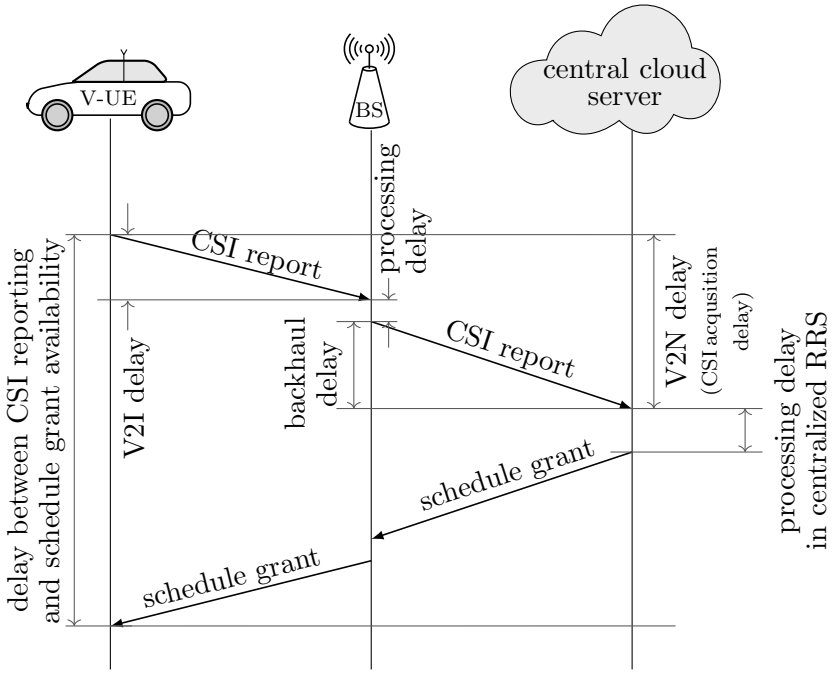


Figure 3.6: Illustration of delay between CSI reporting to the centralized radio resource scheduler and the schedule grant availability at a V-UE.

track QSI for each multicast group across CAD groups considering the periodicity of the respective data traffic. Yet, the CSI acquisition poses the aforementioned challenges. To this end, Sect. 3.3 proposes a centralized dynamic RRS which is robust against the V2N delay and reduces the signaling overhead over V2N links significantly.

3.3 Slow CSI-based Dynamic Resource Scheduling

In order to reduce the signaling overhead over V2N links, the sidelink RRM function can configure the V-UEs to report only averaged CSI measurements at a rate $1/T_{\mathrm{c}}$, where $T_{\mathrm{c}} \gg T_{\mathrm{sl}}$, as opposed to full CSI measurements acquisition at each time slot. The averaged CSI values contain slowly varying components of the channel gain (the large-scale fading components) such as the path loss. Hence, such averaged

measurements are referred to as *slow CSI*. The latter remains almost unchanged for a V2V link over control plane delays of $50 \ldots 100$ ms incurred over V2N links. These delays comprise the time duration from CSI reporting at the V-UEs to the usage of sidelink resources that are scheduled based on the reported CSI for V2V transmission and reception as illustrated in Fig. 3.6. The robustness of slow CSI against the V2N control plane delay can be exploited in RRS for reliable V2V transmissions.

Let $\bar{H}_{\mathrm{L}_{d,\tilde{d}},n}$ denote the latest acquired slow CSI for the V2V link $\mathrm{L}_{d,\tilde{d}} \in \mathcal{L}$ available at the RRS at time slot n. Here, assuming a unit-mean small-scale fading, $\bar{H}_{\mathrm{L}_{d,\tilde{d}},n}$ is identical to $\eta_{\mathrm{L}_{d,\tilde{d}},n}$ and is practically constant over T_{c} slots. Then, for a given allocation $\mathbf{\Pi}$ at time slot n, $\bar{\gamma}_{\mathrm{L}_{d,\tilde{d}},k,n}(\mathbf{\Pi})$ denotes the SINR computed based on only slow CSI and can be obtained by using $\bar{H}_{\mathrm{L}_{d,\tilde{d}},n}$ instead of $H_{\mathrm{L}_{d,\tilde{d}},k,n}$ in (2.3). Unlike the SINR $\gamma_{\mathrm{L}_{d,\tilde{d}},k,n}(\mathbf{\Pi})$ in (2.3), this slow CSI-based SINR $\bar{\gamma}_{\mathrm{L}_{d,\tilde{d}},k,n}(\mathbf{\Pi})$ is influenced by only the large-scale fading components. Hence, it changes sufficiently slowly and remains almost unchanged over T_{c} slots. Subsequently, the slow CSI-based channel quality $\bar{\gamma}_{m,k,n}(\mathbf{\Pi})$ for the m^{th} multicast group transmission at time slot n in the k^{th} RB is defined using the slow CSI-based SINR $\bar{\gamma}_{\mathrm{L}_{d,\tilde{d}},k,n}(\mathbf{\Pi})$ in (2.4).

3.3.1 Problem Description using SINR Margin

The slow CSI cannot be directly used in the centralized RRS such as the solution proposed in Sect. 3.2.2 since it does not contain the information on fast fading components of the channel. The fast fading component influences the SINR performance and may result in outage of the V2V link, wherein the outage refers to the inability of the link to provide the required or anticipated SINR performance. Here, given a slow CSI-based allocation $\mathbf{\Pi}$ with $\Pi_{m,k} \in \Omega_{1,A}$, the outage probability for V2V transmission of the m^{th} multicast group at time slot n in the k^{th} RB reads

$$\Pr\Big(\gamma_{m,k,n}(\mathbf{\Pi}) \leq \gamma_{\mathrm{thr},m}(\Pi_{m,k})\Big). \tag{3.5}$$

The RRS must take into account the fast fading induced outage to schedule reliable V2V transmissions. In particular, the reliability constraint (3.4c) in the resource allocation problem (3.4) needs to be further modified to consider the outage probability (3.5). To this end, considering different reliability requirements, this fast-fading induced outage probability must be constrained to meet a corresponding confidence level α_m for the m^{th} multicast group. Hence, the reliability constraint (3.4c) in the resource allocation problem can be modified as

$$\Pr\Big(\gamma_{m,k,n}(\mathbf{\Pi}) \leq (2\mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) - 1)\gamma_{\text{thr},m}(\Pi_{m,k})\Big) \leq \alpha_m. \tag{3.6}$$

Based on the reliability requirement, an appropriate confidence level α_m must be configured at the sidelink RRM function for each multicast group. Here, the higher the reliability requirements, the lower is the confidence level in (3.6) induced by fast fading. It is noted that (3.6) holds when no resource is allocated to the m^{th} multicast group, i.e., $\Pi_{m,k} = 0$.

Since the probability of the fast-fading induced outage can be reduced by allocating resources in a manner that provides an extra SINR margin [60], this work makes use of the SINR margin for slow CSI-based resource allocation to meet the reliability requirements. This extra SINR margin accounts for the potential deviation of the SINR performance due fast fading which is not represented by the slow CSI measurements. Let ξ_m denote the SINR margin for the m^{th} multicast group transmissions. The higher the SINR margin, the lower is the outage probability. Furthermore, considering the strict reliability requirements of CAD V2V transmissions, let $\xi_{\text{thr},m}(\alpha_m)$ be the minimum SINR margin that must be ensured so that the outage constraint in (3.6) is met for the required confidence level α_m. Then, given such $\xi_{\text{thr},m}(\alpha_m)$, the reliability constraint (3.4c) can be further modified as

$$\big(2\mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) - 1\big)\, \xi_{\text{thr},m}(\alpha_m)\gamma_{\text{thr},m}(\Pi_{m,k}) < \bar{\gamma}_{m,k,n}(\mathbf{\Pi}), \tag{3.7}$$

so that the acquisition of full CSI is avoided and reliability (or SINR) requirements are met based on only slow CSI in the dynamic resource allocation.

An example for a minimum SINR margin is as follows. For simplicity, consider a Rayleigh-Rayleigh environment where both desired and interference signals are subject to Rayleigh fading. Assuming an interference power to be much larger than the thermal noise power, the outage probability can be obtained from [60] as

$$\Pr\Big(\gamma_{m,k,n}(\mathbf{\Pi}) \leq \gamma_{\mathrm{thr},m}(\Pi_{m,k})\Big) \leq 1 - \exp\left(-1/\xi_m\right).$$

Then, for the required α_m with $0 < \alpha_m < 1$, the SINR margin for the m^{th} multicast group is given by

$$\xi_{\mathrm{thr},m}(\alpha_m) = -\frac{1}{\log\left(1-\alpha_m\right)}. \tag{3.8}$$

Given the slow CSI measurements (as oppposed to full CSI) at all V-UEs, the resource allocation problem considers to provide an extra SINR margin for V2V transmissions at each multicast group to counter the fast-fading induced outage in order to meet the reliability requirements. Hence, modifying (3.4) for slow CSI-based scheduling, the resource allocation problem can be reformulated as

$$\grave{\mathbf{\Pi}}_{\mathrm{opt},n} = \arg \max_{\mathbf{\Pi} \in \Omega_{0,A+1}^{M \times K}} \sum_{m \in \mathcal{F}_n} u_{\mathrm{del},m}\left(\tau_{m,n}(\mathbf{\Pi})\right) \tag{3.9}$$

subject to

$$\Pi_{m,k} = 0 \quad \forall\, (m,k) \in \Omega_{1,M} \backslash\, \mathcal{F}_n \times \Omega_{1,K} \tag{3.9a}$$

$$\sum_{m \in \mathcal{M}_{\mathrm{CAD},x}} \max_{k \in \Omega_{1,K}} \mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) \leq 1 \quad \forall x \in \Omega_{1,X} \tag{3.9b}$$

$$\left(2\mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) - 1\right) \xi_{\mathrm{thr},m}(\alpha_m) \gamma_{\mathrm{thr},m}(\Pi_{m,k}) < \bar{\gamma}_{m,k,n}(\mathbf{\Pi}) \;\; \forall (m,k) \in \mathcal{F}_n \times \Omega_{1,K}. \tag{3.9c}$$

Here, the constraints (3.9a) and (3.9b) are identical to (3.4a) and (3.4b), resp.,

to restrict the resource allocation exclusively to multicast groups with non-empty queues and to atmost one multicast group within a CAD group at any time slot considering the half-duplex constraint, respectively. Now (3.9c) modifies (3.4c) as per (3.7) to assure that the reliability requirements of multicast groups are satisfied by providing an extra SINR margin to meet (3.3).

3.3.2 Solution

Similar to (3.4), the problem (3.9) is a constrained optimization problem, and therefore, to provide a solution with low complexity for a simpler implementation, an iterative algorithm is presented in **Algorithm** 3 for resource allocation. As in Sect. 3.2.2, the scheme makes use of **Algorithm** 1 to reduce the solution search space by nominating at most one multicast group for resource allocation from each CAD group. It is noted that **Algorithm** 1 nominates at most one multicast group with non-empty queues in each CAD group which is most preferred (i.e., the multicast group that returns the maximum utility) for resource allocation at the given time slot. Therefore, by considering only those multicast groups that are nominated for resource allocation by **Algorithm** 1, the constraints (3.9a) and (3.9b) can be satisfied.

Similar to **Algorithm** 2, **Algorithm** 3 iteratively assigns RBs and MCSs to the nominated multicast groups in $\check{\mathcal{F}}_n$ by selecting multicast groups for allocation in each RB based on a delay-dependent utility-based greedy criterion (i.e., the negative slope of the utility function for the HOL delay at the end of the current time slot). Moreover, the algorithm incorporates the extra SINR margin in line 9 and line 18 to assure that the reliability (SINR performance) requirements of multicast groups are met in spite of slow CSI usage for resource allocation. It is noted that **Algorithm** 3 differs from **Algorithm** 2 mainly in line 9 and line 18 as it introduces the extra SINR margin and makes use of the slow CSI-based channel quality metric.

Algorithm 3: Slow CSI-based sidelink RRS Algorithm

Input : slow CSI and QSI with $\check{\mathcal{F}}_n$

Output: $\mathbf{\Pi}$

Initialization: $\mathbf{\Pi} \leftarrow \mathbf{0}_{M\times K}$;
for $k = 1 : K$ **do**
  $\mathcal{M} \leftarrow \check{\mathcal{F}}_n$;
  while $|\mathcal{M}| > 0$ **do**
    $\acute{m} \leftarrow \arg\max_{\tilde{m}\in\mathcal{M}} -u'_{\mathrm{del},\tilde{m}}\left(\tau_{\tilde{m},n}(\mathbf{\Pi})\right)$;
    for $a = A : 1$ **do**
      $\tilde{\mathbf{\Pi}} \leftarrow \mathbf{\Pi}$;
      $\tilde{\Pi}_{\tilde{m},k} \leftarrow a$;
      if $\xi_{\mathrm{thr},\acute{m}}(\alpha_{\acute{m}})\gamma_{\mathrm{thr},\acute{m}}(\tilde{\Pi}_{\tilde{m},k}) - \bar{\gamma}_{\acute{m},k,n}(\tilde{\mathbf{\Pi}}) \leq 0$ **then**
        $\Pi_{\acute{m},k} \leftarrow \tilde{\Pi}_{\tilde{m},k}$;
        break;
      end if
    end for
    $\tilde{\mathcal{M}} \leftarrow \mathcal{M} \setminus \{\acute{m}\}$;
    foreach $\tilde{m} \in \mathcal{M}$ **do**
      $\tilde{\mathbf{\Pi}} \leftarrow \mathbf{\Pi}$;
      $\tilde{\Pi}_{\tilde{m},k} \leftarrow 1$;
      if not $\bar{\gamma}_{m,k,n}(\tilde{\mathbf{\Pi}}) > (2\mathbb{1}_{\Omega_{1,A}}(\tilde{\Pi}_{m,k}) - 1)\xi_{\mathrm{thr},m}(\alpha_m)\gamma_{\mathrm{thr},m}(\tilde{\Pi}_{m,k})$ $\forall m \in \mathcal{M}$ **then**
        $\tilde{\mathcal{M}} \leftarrow \tilde{\mathcal{M}} \setminus \{\tilde{m}\}$;
      end if
    end foreach
    $\mathcal{M} \leftarrow \tilde{\mathcal{M}}$;
  end while
end for

3.3.3 Advantages and Challenges

Finding the solution to the slow CSI based resource allocation problem (3.9) needs a centralized dynamic scheduler and requires the placement of sidelink dynamic RRM

functionalities at the central cloud server. In addition, it requires slow CSI measurements for appropriate MCS selection and instantaneous QSI for delay calculation and identifying non-empty queues. Being a centralized dynamic resource allocation solution, it inherits the advantages and challenges of the instantaneous QSI-based centralized dynamic resource allocation scheme discussed in Sect. 3.2.2, yet with the relaxation that each V-UE must report only the slow CSI instead of the full CSI for the associated V2V links.

The proposed solution significantly reduces the signaling overhead since the V-UEs need to report only the slow CSI measurements at a rate $1/T_{\mathrm{c}}$ as opposed to full CSI measurements acquisition at each time slot. This is particularly beneficial in reducing the operation cost for sidelink RRM when the sidelink RRM functionality is using the V2N links as *pay-per-use* service offered by MNOs that are managing the V2N network infrastructure. Also, the resource allocation is robust against the V2N delay (as slow CSI remains almost unchanged over V2N delays) and ensures reliable V2V transmissions. Besides, it inherits advantages from the full CSI-based centralized solution (described in Sect. 3.2.2) such as its ability to adapt to diverse resource demands from CAD operations by means of utility-based RRS, handling the half-duplex constraint and the possibility to decouple from MNOs' networks.

While the extra SINR margin enables the solution to ensure high reliability, the sidelink resource utilization is reduced as the reuse of the RBs may be limited due to increased SINR margins (i.e., the number of multicast groups that are allowed to use the same RB may be reduced). This may adversly impact the delay performance across the multicast groups. Furthermore, inheriting from the centralized cloud-enabled RRS, the V2N delay could be significantly larger depending on the deployed backhaul which may result in a higher scheduling delay and therefore a larger queuing delay. This could potentially lead to sub-optimal CAD operation and may not be acceptable for critical CAD operations such as CCA. This is particularly

challenging for NPDT that cannot be predicted with high certainty. In case of PDT (or highly predictable) V2V transmissions, however, this challenge can be overcome by implementing a virtual queue in the RRM function that can track QSI for each multicast group (e.g., based on the periodicity of the PDT) and perform resource allocation based on the virtual queue beforehand without requiring the explicit QSI from V-UEs. Therefore, in case of a large V2N delay, the solution is more suitable for PDT V2V communications.

Chapter 4

Edge Cloud-Enabled Semi-Centralized Sidelink Radio Resource Management for CAD

Cloud-enabled sidelink RRM is proposed in Chapter 3 which allows for centralized QoS-aware sidelink RRM in an operator-independent manner to support CAD. However, this necessitates the placement of a dynamic sidelink radio resource scheduler in a logically centralized cloud server, which requires all the RRC information to traverse through V2N links with potentially high transport network delays. This, as identified in Sect. 3.2.3, entails higher scheduling delays as well as signalling overhead over V2N links which comprise the air interfaces between V-UEs and BSs (i.e., V2I links) as key challenges. To this end, while Chapter 3 considers robust schemes for central (remote) cloud-based RRS, this chapter proposes an approach to bring the scheduler closer to the V-UEs to decrease the scheduling delay and signaling overhead while exploiting the benefits of cloud computing for sidelink RRM.

To reduce the latency between the cloud server and the end users, edge computing has been introduced [18]. In this paradigm, the computing resources are made avail-

able at the edge of the network near to or co-located with the end devices [19]. When implemented for vehicular scenarios, this can provide a one-hop-away low-latency link between the edge cloud and the vehicles, thereby reducing the data shared with distant central cloud entities (data centers) and allowing real-time access to radio link information at the edge cloud. Furthermore, this decentralized geographically distributed cloud platform supports mobility of vehicles. Therefore, utilizing both edge and central cloud computing for sidelink RRM, a hybrid cloud framework can offer flexibility in placing RRM functionalities in edge and central cloud entities. This framework combines the benefits from centralized resource control (e.g., interference coordination across MNOs) and one-hop distributed scheduling (e.g., low scheduling delay).

4.1 Radio Resource Management Framework

The high dynamics of vehicular environments due to the mobility of vehicles in CAD groups are a key challenge in CAD use cases. This may strongly impact the sidelink performance on both user and control planes in fully distributed or fully centralized implementations of sidelink RRM functionalities. While a fully centralized implementation enables QoS-aware sidelink RRC across all V-UEs allowing for radio resource reuse, it may introduce increased delay over the V2N links between the central (or remote) cloud and V-UEs because of high transport network delays. Due to this, the acquired information such as CSI in the resource scheduling function may become outdated (irrelevant) resulting in unreliable scheduled transmissions. In addition, this higher delay may increase the scheduling delay which could potentially lead to sub-optimal CAD performance and can be critical in emergency scenarios. Similarly, a large degree of centralization results in an increased signaling overhead over V2N links in dynamic vehicular environments. On the other hand, a fully distributed implementation may reduce the sidelink channel access de-

lay by allowing coordination among neighboring V-UEs for sidelink transmissions. However, this may result in lower reliability performance on V2V links due to an increased number of hidden nodes as well as packet loss due to half-duplex limitations in high-mobility environments. Therefore, a semi-centralized implementation of sidelink RRM that exploits the benefits of both centralized and distributed implementations can benefit CAD use cases in providing low-latency and high reliability under high vehicular mobility. Furthermore, a suitable cloud platform is required to host such semi-centralized sidelink RRM functions to support decoupling of sidelink RRM from operators' networks. To this end, the cloud-enabled sidelink RRM framework is extended by leveraging the VEC capabilities of autonomous vehicles to allow semi-centralized implementation of sidelink RRM.

In the following, Sect. 4.1.1 introduces a hybrid edge cloud platform where the edge computing can be implemented as an edge cloud federation of cloud and VEC nodes to exploit the benefits of both centralized and distributed computing. Here, remote cloud nodes provide logically centralized cloud computing capabilities to a set of BSs across MNOs. At the same time, leveraging the on-board computational capabilities of the autonomous vehicles, VEC exploits/adds computational resources from neighboring or co-located vehicles. In this hybrid edge cloud implementation, using the cellular coverage of the network, cloud nodes can configure VEC nodes to perform functional computation. Then, Sect. 4.1.2 splits the sidelink RRM into three modular RRM functionalities (or tasks) which could be placed flexibly in different edge cloud entities to address the trade-off among multiple performance indicators such as control plane delay, network throughput, reliability and signaling overhead.

4.1.1 Edge Cloud Infrastructure

The autonomous vehicles are famously referred to as *computer on wheels* [61] owing to their computational capability which is primarily used to support different LoAs.

The on-board data storage and computational resources of autonomous vehicles can be exploited to make cloud computing capability available at nearby or co-located vehicles. In this regard, VEC has been introduced in [31], [32] where an autonomous vehicle can be viewed as a mobile (or moving) infrastructure at the edge of the network which offers data storage and computational resources as a VEC node. In addition to the on-board VEC resources, a vehicle can leverage its V2V communication capability to utilize the resources of neighboring VEC nodes (i.e., vehicles offering VEC) for either offloading computation or to add additional computation capabilities. That is, the vehicle may share the available VEC resources via V2V communications. Fig. 4.1 illustrates the interaction among VEC nodes where a vehicle can either offload its computation task to the one-hop-away neighboring vehicle with VEC resources or share its VEC resources with the one-hop-away neighboring vehicle. It is noted that VEC is also known as vehicular fog computing [31].

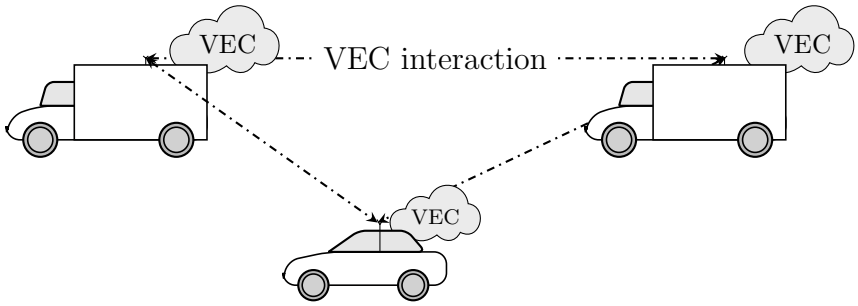


Figure 4.1: Illustration of VEC node interaction.

Because of the ability to provide cloud computation capability at a one-hop-away VEC node, VEC enables vehicles to access the cloud computing resources with low latency. Also, it eliminates the need to share the data with the distant central cloud entities (data centers) for cloud computing. Furthermore, the VEC nodes are decentralized, geographically distributed and mobile. Hence, when vehicles are travelling in groups (e.g., as platoons), one or more vehicles in each group can act as a VEC node to provide cloud computation capability to all the vehicles within the group.

To benefit from remote cloud servers[1] offering logically centralized cloud computing capability for all the vehicles under its coverage, remote clouds are considered with VEC nodes to provide an edge cloud hybrid computing infrastructure as shown in Fig. 4.2. Here, V2N links can be used to offload computation tasks from VEC

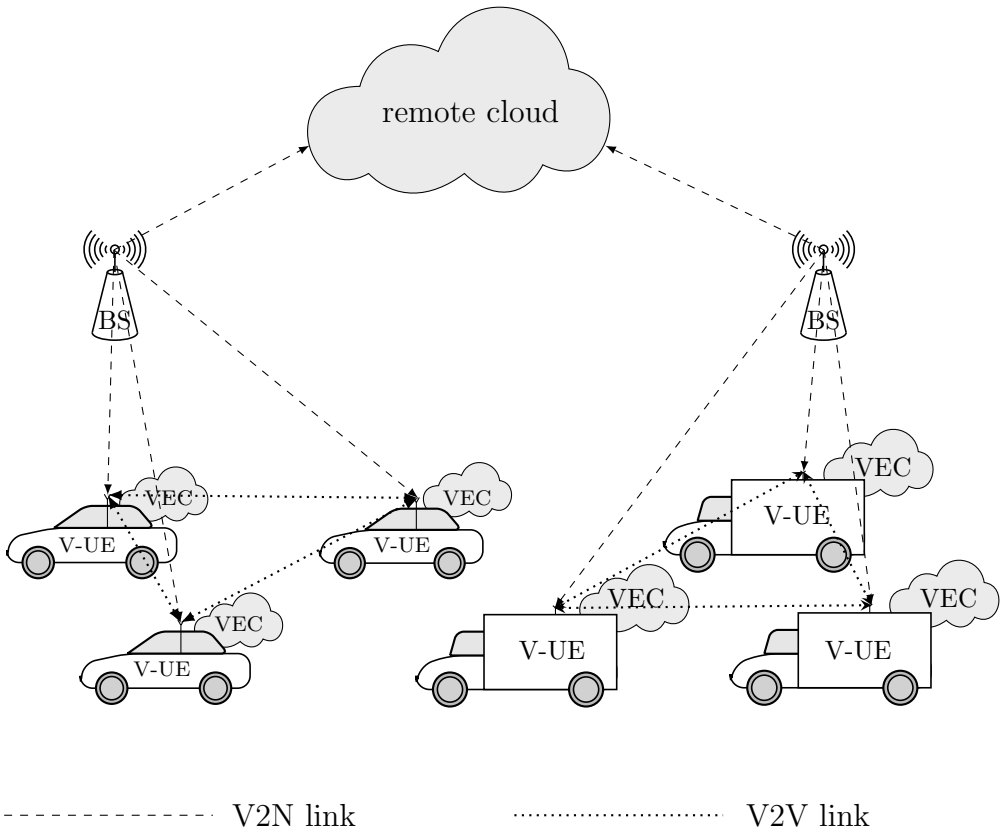


Figure 4.2: Illustration of hybrid cloud computing platform.

nodes to remote clouds or vice versa and for sharing cloud computing resources between VEC nodes and remote clouds. V2V links are used to offload computation tasks and share cloud computing resources among VEC nodes.

Creation and maintenance of a hybrid computing platform as federation of remote clouds and VEC nodes have been studied recently [38], [39], where [38] introduces a HVC framework for hybrid computing. In this work, we exploit a hybrid computing platform approach for sidelink RRM, but do not deal with the creation

[1]Clearly, the subsequent treatment also holds in case of a single cloud server.

and/or maintenance of such a platform.

4.1.2 RRM Functional Split

To support semi-centralized implementation of sidelink RRM to selectively exploit the benefits of both centralized implementation such as interference management and hence higher reliability and of distributed implementation such as low latency scheduling while aiming to reduce signaling overhead over V2N, the sidelink RRM is split into three modular functions (tasks). This splitting of tasks allows to utilize the hybrid cloud computation platform to flexibility carry out sidelink RRM tasks at the centralized cloud node (e.g., remote cloud server), at the VEC node or both. Furthermore, the approach allows for interaction between the tasks via V2N links on different time scales to address the trade-off among multiple performance indicators such as packet delay (dependent on scheduling delay), network throughput, reliability over V2V links and signaling overhead over V2N links. The proposed modular sidelink RRM functions with mutual interaction are illustrated in Fig. 4.3 and comprise the following tasks.

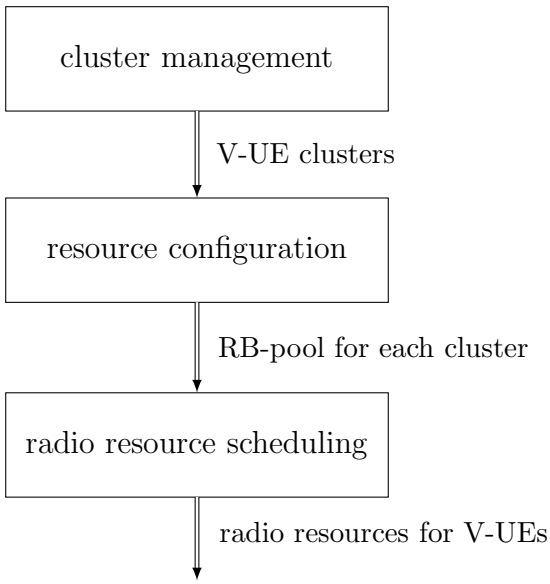


Figure 4.3: Graphical representation of sidelink RRM functional split.

- Cluster management (CM): An overarching multi-cell V-UE clustering function that forms V-UE clusters based on operational policies (e.g., cell coverage-based, geographical area-based resource allocation) and considering application requirements on V-UE grouping. The CM function outputs V-UE clusters.

- Resource configuration (RC): An RRC functionality that configures an RB-pool (consisting of a set of RBs) to V-UE clusters formed by the CM function for unicast and multicast transmissions. Here, V-UEs in the clusters are restricted to use the resources exclusively from the assigned RB-pool for V2V communications. Furthermore, inter-cluster interference is ensured to be limited to an acceptable level. The RC function takes V-UE clusters as input and outputs an RB-pool to each V-UE cluster.

- RRS: A MAC functionality for dynamically allocating RBs in the RB-pool to the V-UEs based on the target resource allocation objective. The RRS function takes an RB-pool assigned to each V-UE cluster as input and outputs radio resources to one or more V-UEs in each cluster from the respective RB-pool.

4.1.3 RRM Function Placement in Edge Cloud Entities

Fig. 4.4 illustrates different example implementations of sidelink RRM on edge cloud infrastructures that exploit the aforementioned functional split to achieve different objective (e.g., low signaling overhead over V2N links, high reliability and/or low latency on V2V links, etc.) and to serve different CAD use cases. Also, the clustering (e.g., proximity-based, location-based, link quality-based), the RB-pool assignment (e.g., distance-based, interference-based, QoS-aware) and dynamic RRS (e.g., centralized or distributed QoS-aware, channel sensing-based) strategies are chosen accordingly.

Implementation-A (Fig. 4.4a) places CM and RC functions at the central cloud

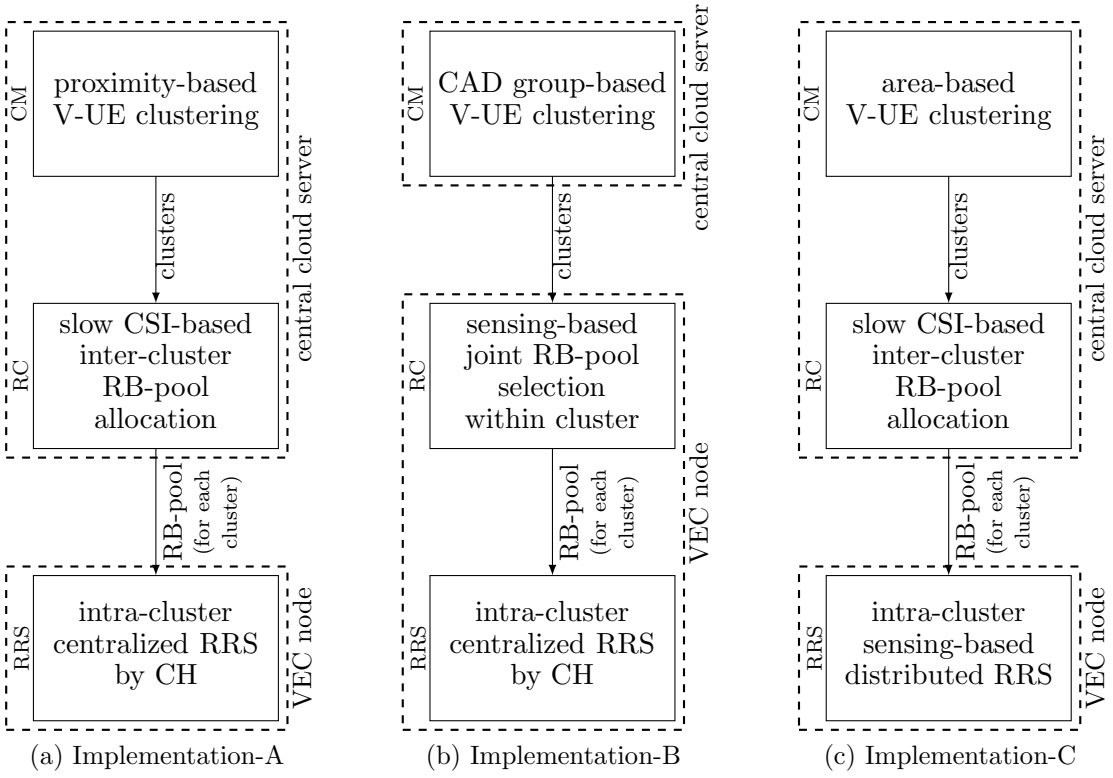


Figure 4.4: Graphical representation of sidelink RRM functional split examples.

entity and the RRS function at the VEC node to aim for a low scheduling delay and a high reliability. Here, nearby V-UEs are clustered at the central cloud server so that the cluster head (CH) V-UE can perform dynamic resource allocation at vehicle level within the cluster to reduce the scheduling delay. In addition, an RB-pool is administrated centrally at the cloud server to increase the radio resource reuse while ensuring the inter-cluster interference to stay within an acceptable limit for increased reliability. Moreover, a slow CSI (as introduced in Sect. 3.3) can be used at the RC function to reduce signaling overhead over V2N links, where V-UEs can be configured to report the measurement periodically at an interval of hundreds of milliseconds.

Implementation-B (Fig. 4.4b) places only the CM function at the central cloud entity and considers each CAD group as a cluster to significantly reduce the signaling

overhead over V2N links. At the same time, the RC and RRS functions are placed at the VEC nodes where the RC function can configure V-UEs within the cluster to perform joint RB-pool selection based on channel-sensing schemes such as LTE-V Mode-4 or NR V2X Mode-2 as proposed in [47], [48]. However, this increases the signaling overhead over V2V links in determining the RB-pool for the V-UE cluster and the radio resource utilization may be degraded as the optimal RB-pool assignment to maximize the resource reuse may not be achieved in a distributed RB-pool assignment across the clusters. The intra-cluster allocation can be similar to that of Implementation-A. Also, this is most suitable when all or most of the V-UEs are involved in PDT transmission as the V-UE can predict the interference over V2V more accurately and hence can perform RRS accordingly to achieve higher reliability.

Implementation-C (Fig. 4.4c) places CM and RC functions at the central cloud entity while placing the RRS function at the VEC node and performs geographical area-dependent RRS to provide low latency on V2V links and to significantly reduce signaling overhead over V2N links. The geographical area can be divided into grid-like zones and all the V-UEs that are located in a zone are considered to be a cluster. For RB-pool allocation, the RC function can be configured to collect and store location-dependent slow CSI measurements (including both transmitter and receiver V-UE location information) and report periodically (e.g., once in tens or hundreds of milliseconds) to the RC function. From the collected measurements, the RB-pool is updated centrally at the cloud server to increase the radio resource reuse while reducing the hidden nodes and ensuring the inter-cluster (i.e., inter-zone) interference to within an acceptable limit for increased reliability as proposed in [49]. For intra-cluster resource allocation, the V-UEs can be made to perform sensing-based (e.g., LTE-V Mode-4 or NR V2X Mode-2) schemes for resource selection. However, the resource utilization may be degraded when resource demands vary highly dy-

namically. Also, the reliability performance may be degraded as the V-UEs do not coordinate their transmissions to avoid packet loss due to the half-duplex constraint. As a special case, when the zone is chosen to be the radio coverage of a vehicle, a V-UE can be made to select all the RBs in the associated RB-pool when located in a zone with no scheduling delay as proposed in [50]. This is particularly beneficial for safety critical CAD use cases such as CCA. However, such pre-allocation may significantly impact the resource utilization as RBs assigned for unoccupied zones might be wasted. Yet, it might be suitable in specific areas with high safety concerns.

It is noted that the centralized scheduling can be realized by placing all the three functions in a centralized cloud entity where the CM function considers all the V-UEs to be a part of a single cluster and all RBs are considered to be in the associated RB-pool by the RC function. Then, a single RRS function can perform dynamic resource allocation for all V-UEs. On the other hand, a fully distributed scheduling can be realized by placing all the three functions in VEC nodes where a CM function can cluster V-UEs using, e.g., request-response based signaling exchange between V-UEs (an example is described in [47]). The RC function can then configure V-UEs to jointly perform an RB-pool selection using channel-sensing based schemes. One such possibility to perform a joint RB-pool selection is described in [47]. Moreover, while this work considers the federation of a central cloud and VEC nodes, the proposed framework may also be applicable to different edge cloud realizations where the V-UE functionalities could be placed flexibly in different edge cloud entities subject to the edge cloud deployment such as the federation of remote clouds, MEC, cloudlets and/or VEC nodes.

In this work, a semi-centralized scheme following essentially implementation-A is treated considering its potential to support diverse QoS requirements of CAD applications including low latency and high reliability over V2V links.

4.2 Three-Stage Resource Scheduling

In the proposed implementation framework, the sidelink RRM functionality is split into long-term RRM functionalities, i.e., CM and RC, and a dynamic RRS functionality exploiting the edge cloud infrastructure. CM and RC functions update V-UE clusters and the RB-pool configuration, resp., periodically. The periodicity of V-UE clustering and RB-pool configuration may vary depending on the deployment and vehicular environments. For example, for urban scenarios with usually low mobility, the update is less frequent to reduce the signaling overhead, while in scenarios involving high mobility of vehicles, e.g., motorways, the clustering and RB-pool configuration will be done more frequently. On the other hand, the resource scheduler is placed at the VEC node formed by the vehicles in the cluster. The output (clusters and RB-pool) of CM and RC functionalities are conveyed to the RRS functions through the V2N links. Due to the close proximity of the dynamic resource scheduler placed at VEC nodes within each moving cluster, the scheduling delay is significantly reduced and the scheduler can acquire full CSI within each cluster over one-hop away V2V links for V2V link adaptation.

The framework considers V2X application layer aspects described in Sect. 3.1.2 including an V2X application server and its functionalities and interaction with V-UEs. Also, this framework places the CM and RC functions in the V2X application server to allow interaction with the V2X application. In addition, the interaction between the sidelink RRS and V2X application server can be enabled via CM and RC functions or via a direct V2N link for QoS-aware resource allocation.

Fig. 4.5 illustrates the proposed sidelink RRM functional split implementation on the hybrid cloud computing platform. In 3GPP specified 5G networks, the application layer functional support to V2X services offered by the 3GPP can be exploited to realize the functional split as shown in Fig. B.2 of Appendix B. By exploiting this

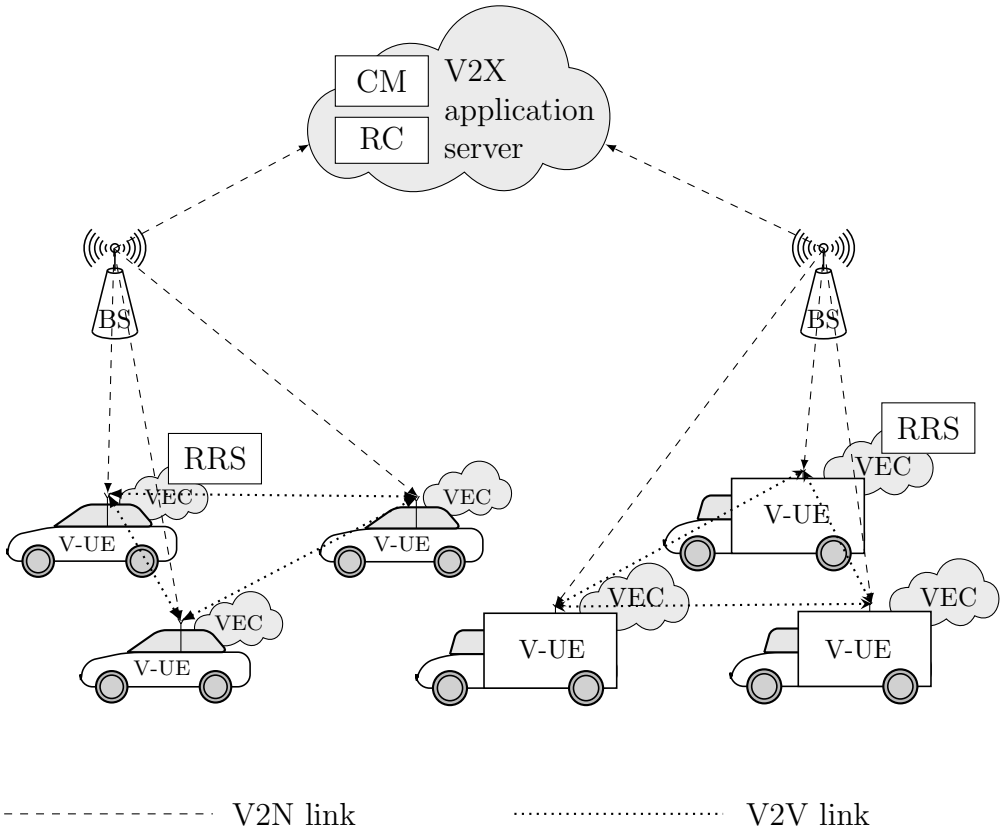


Figure 4.5: Placement of sidelink RRM functions on a hybrid cloud computing platform.

framework, the problem of resource allocation for CAD V2V multicast transmission is tackled in three stages as described in Sect. 4.2.1. A graph-theoretic approach to solve the problem is presented in Sect. 4.2.2.

4.2.1 Problem Description

As described in Sect. 3.2.1, this work models the CAD QoS requirements in terms of utility functions and views the resource allocation for CAD V2V communications an optimization problem to maximize the aggregated QoS-dependent utility across multicast groups. However this optimization across all multicast groups requires a global (covering all V-UEs) centralized dynamic scheduler and hence necessitates placement of the scheduler at the central cloud server. Therefore, the approach

suffers from a usually substantial RRC signaling delay and high signaling load over V2N links. To this end, the problem is relaxed to allow scheduling for a subset of multicast groups locally (instead of globally) at vehicle level at a VEC node. However, this relaxation may impact the performance in terms of resource utilization and QoS on V2V links depending on the resource availability at each local scheduler, the level of interference caused at a multicast group by the RRS decisions of other local schedulers, etc. Therefore, choosing a subset of multicast groups for local scheduling and the resources that are utilized in each local scheduler play a key role in minimizing the impact of relaxation. In this regard, the resource allocation problem is split into three stages as illustrated in Fig. 4.6, which offer high modularity

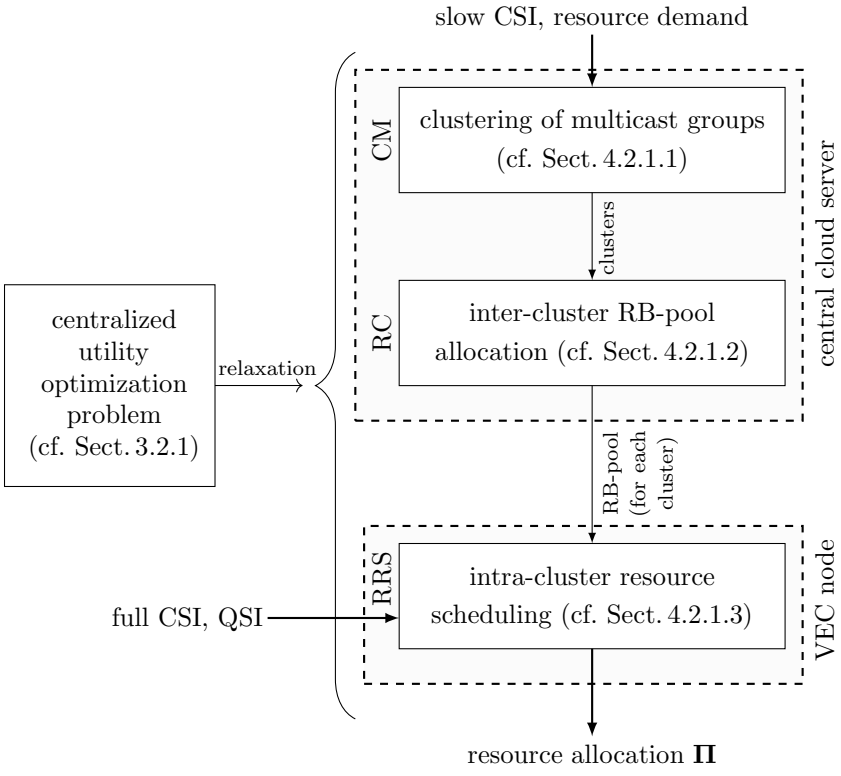


Figure 4.6: Graphical representation of three-stage resource allocation problem.

and tunability, and can be performed at different sidelink RRM functions as per the proposed edge cloud-enabled sidelink RRM framework. The three stages are used

to solve the following tasks:

1) a multicast group cluster formation in a CM function,

2) an inter-cluster RB-pool allocation in an RC function,

3) an intra-cluster resource scheduling in an RRS function.

Here, the multicast group cluster formation task involves clustering of multicast groups to allow for transmission and resource co-ordination among nearby V-UEs and ensures control plane reachability on V2V links among the transmitting V-UEs of multicast groups within the cluster. This co-ordination among V-UEs avoids packet loss due to half-duplex limitations and the hidden-node problem when the interference-limited RBs are made available at each cluster. In this regard, the inter-cluster RB-pool task assigns to each cluster a pool of interference-limited RBs from which RBs can be selected for V2V transmissions within each cluster. In addition, the inter-cluster RB-pool allocation provides a trade-off between resource utilization and V2V link quality in clusters by controlled spatial reuse of RBs limiting the inter-cluster interference to an acceptable level. Upon clustering of multicast groups and inter-cluster RB-pool allocation, an intra-cluster resource allocation allows for dynamic allocation of RBs from the configured RB-pool and MCSs to multicast groups within each cluster. Since control plane reachability among the transmitting V-UEs of multicast groups is ensured while clustering of multicast groups, transmitting V-UEs within the cluster can perform dynamic scheduling either in a distributed fashion or by electing a CH and provide schedule grants via V2V links with reduced scheduling delay and hence eliminating the needs for V2N links for dynamic scheduling.

In the proposed approach, cluster formation and inter-cluster RB-pool allocation are performed at the cloud server in CM and RC functions, resp., as long-term RRM

tasks. Here, the cloud server (as a long-term scheduling entity) acquires measurements of only the slow CSI and resource demand estimates at a rate $1/T_{\mathrm{c}}$, where $T_{\mathrm{c}} \in \mathbb{N}$ denotes the time slots for clustering multicast groups for the next interval. In the following, clustering and RB-pool allocation is considered for an arbitrary interval, and hence the notations related to the variable do not include an index for the time slot for convenience. Let $\bar{H}_{\mathrm{L}_{d,\tilde{d}}}$ denote the slow CSI over the V2V link $\mathrm{L}_{d,\tilde{d}}$ and $\hat{\Upsilon}_m$ the estimated resource demand for the m^{th} multicast group, respectively. Consequently, the signaling overhead over the V2N link is significantly reduced as opposed to full CSI and QSI acquisition. Furthermore, in order to define bounds on the relative control plane reachability among and data plane interference within, resp., multicast groups, two thresholds $\bar{\psi}_{\mathrm{ctrl}}$ and $\bar{\psi}_{\mathrm{I}}$ are estimated based on slow CSI measurements. Here, $\bar{\psi}_{\mathrm{ctrl}}$ is a threshold for the received power on the control V2V link below which multicast groups cannot coordinate their transmissions with each other, whereas $\bar{\psi}_{\mathrm{I}}$ tackles the trade-off between enhancing spectrum utilization by efficient reuse of sidelink resources among clusters and limiting the inter-cluster interference to an acceptable level. This offers high tunability since by adding a positive offset to $\bar{\psi}_{\mathrm{ctrl}}$ the size of the clusters can be influenced. Here, the higher $\bar{\psi}_{\mathrm{ctrl}}$, the smaller is the cluster size. Also, by varying $\bar{\psi}_{\mathrm{I}}$ the resource reuse and hence the inter-cluster interference can be influenced, and the higher $\bar{\psi}_{\mathrm{I}}$, the higher is the resource reuse, but the lower is the inter-cluster interference. On the other hand, the intra-cluster optimum dynamic resource allocation can be performed at the vehicle level in the resource scheduling function at the VEC node by sharing full CSI and QSI locally within the cluster.

4.2.1.1 Long-term Clustering of Multicast Groups

At the first stage, multicast groups are clustered to allow for channel access coordination among the transmitters of multicast groups within the cluster. The access

coordination involves transmission coordination to avoid reception failure due to half-duplex constraints and intra-cluster RRS to maximize the sum of individual utilities of multicast groups. Therefore, control plane reachability among the transmitters of the multicast groups in the cluster is necessary.

To account for the ability of the transmitter of a multicast group to reach a corresponding node of another multicast group, a control plane reachability matrix $\mathbf{Z} = \{Z_{m,m'}\} \in \mathbb{R}_0^{+M^2}$ is computed where

$$Z_{m,m'} = \begin{cases} P_{\text{ctrl}} \bar{H}_{\text{L}_{d_{\text{tx},m},d_{\text{tx},m'}}} & \text{if } m \neq m' \\ P_{\text{ctrl}} & \text{otherwise.} \end{cases}$$

As a result, if $Z_{m,m'} < \bar{\psi}_{\text{ctrl}}$, the m^{th} and m'^{th} multicast groups must not be in the same cluster as the corresponding transmitting V-UEs cannot coordinate their transmissions with each other. Furthermore, preference or affinity between two multicast groups of indices m and m' to be in the same cluster is defined in terms of a weight $W_{m,m'} \in \mathbb{R}_0^+$ to cluster together multicast groups that have higher affinity with each other. This work defines the affinity between two multicast groups as the channel quality between them on the control plane and chooses $W_{m,m'} = Z_{m,m'}$, and $\mathbf{W} = \{W_{m,m'}\} \in \mathbb{R}_0^{+M^2}$ to be the inter-multicast group weight matrix, while other weights may be chosen depending on the system specifications.

Define $\mathbf{M} = \{M_{m,m'}\} \in (0,1)^{M^2}$ to be a multicast group cluster matrix where

$$M_{m,m'} = \begin{cases} 1 & \text{if the } m^{\text{th}} \text{ and } m'^{\text{th}} \text{ multicast groups are member of the} \\ & \text{same cluster} \\ 0 & \text{otherwise.} \end{cases}$$

Then, in an attempt to cluster multicast groups having a large affinity with each

other, the multicast group clustering problem can be formulated as

$$\hat{\mathbf{M}}_{\text{opt}} = \arg \max_{\mathbf{M} \in (0,1)^{M^2}} \sum_{m \in \Omega_{1,M}} \sum_{m' \in \Omega_{1,M}} M_{m,m'} W_{m,m'} \tag{4.1}$$

subject to

$$M_{m,m'} Z_{m,m'} - M_{m,m'} \bar{\psi}_{\text{ctrl}} \geq 0 \;\; \forall (m, m') \in \Omega_{1,M}^2 \tag{4.1a}$$

$$M_{m,m'} - \prod_{\tilde{m} \in \Omega_{1,M} : M_{m',\tilde{m}} = 1} M_{m,\tilde{m}} = 0 \;\; \forall (m, m') \in \Omega_{1,M}^2 \tag{4.1b}$$

$$M_{m,m'} = 1 \;\; \forall (m, m') \in \mathcal{M}_{\text{CAD},x}^2, \forall x \in \Omega_{1,X}. \tag{4.1c}$$

The constraint in (4.1a) ensures that the multicast groups that cannot perform resource or transmission coordination (i.e., the corresponding link quality is not sufficient for control plane reachability) are not in the same cluster. (4.1b) ensures that all the multicast groups that are clustered together have control plane reachability among each other, which allows for either distributed resource scheduling or joint selection of a CH in case centralized scheduling is preferred within the cluster. (4.1b) implicitly ensures that a multicast group belongs to only one cluster (clustered with only one set of multicast groups). Furthermore, (4.1c) assures that the multicast groups that belong to the same CAD group are clustered together so that they are able to coordinate their transmissions to overcome the limitations resulting from the half-duplex constraint. It is assumed that all multicast groups that are part of a CAD group can reach each other on the control plane as they have been able to carry out CAD operation within a CAD group.

The total number $C \in \mathbb{N}$ of formed clusters equals the rank of $\mathbf{M}$. Furthermore, the set of column indices in each row of the matrix for which the corresponding matrix element is 1 gives a cluster.

4.2.1.2 Long-term Inter-Cluster RB-Pool Allocation

For the further treatment of the RB-pool allocation, let $\mathcal{M}_{\text{clust},c} \subseteq \Omega_{1,M}$ denote the index set of all multicast groups being members of the c^{th} cluster with $c \in \Omega_{1,C}$. Having clustered multicast groups, the RB-pool is allocated to each cluster taking into account the relative resource demand of the involved multicast groups. As the CAD communications are group-based and recipients are in close proximity, RBs can be spatially reused to increase the radio resource utilization. However, the interference at the recipient of any multicast group that may result due to spatial reuse must be limited to an acceptable level. Since the channel quality for a multicast group is limited by the V2V link of minimum channel quality in the group, the maximum interference caused at the respective recipient V-UE set must be considered as the interference at the group. In this regard, to account for the interference at the m^{th} multicast group caused by the m'^{th} multicast group, a slow CSI-based interference metric $\bar{\iota}_{m,m'}$ is defined based on slow CSI as

$$\bar{\iota}_{m,m'} = P_{\text{data}} \max_{d \in \mathcal{D}_{\text{rx},m}} \{\bar{H}_{\text{L}_{d_{\text{tx},m'},d}}\}.$$

Here, the larger P_{data}, the lower is the RB reuse, but the higher is the data signal strength. Furthermore, as the transmitters of multicast groups across clusters do not perform transmission and resource coordination, the worst-case inter-cluster interference must be taken into account while performing inter-cluster RB-pool allocation. To this end, the slow CSI-based worst-case inter-cluster interference metric between the c^{th} and c'^{th} clusters is defined as

$$\bar{\iota}_{\text{clust},c,c'} = \max_{m \in \mathcal{M}_{\text{clust},c}, m' \in \mathcal{M}_{\text{clust},c'}} \bar{\iota}_{m,m'}.$$

Let $\mathbf{Y}$ be a $(C \times K)$-dimensional $(0,1)$-RB-pool matrix with

$$Y_{c,k} = \begin{cases} 1 & \text{if the } k^{\text{th}} \text{ RB is in the RB-pool of the } c^{\text{th}} \text{ cluster} \\ 0 & \text{otherwise.} \end{cases}$$

Then, the number of RBs allocated to the c^{th} cluster is given by

$$y_c = \sum_{k \in \Omega_{1,K}} Y_{c,k}.$$

Furthermore, to account for the proportionality of the RB-pool size y_c assigned to the aggregated resource demand of the multicast groups within the c^{th} cluster, a measure $f_c = y_c / \Gamma_c$ is defined as the RB-pool allocation outcome with

$$\Gamma_c = \left(\min_{m \in \Omega_{1,M}} \hat{\Upsilon}_m \right)^{-1} \sum_{m \in \mathcal{M}_{\text{clust},c}} \hat{\Upsilon}_m,$$

where it is assumed that $\min_{m \in \Omega_{1,M}} \hat{\Upsilon}_m > 0$. Then, given an RB-pool size vector $\mathbf{y} = [y_1, \cdots, y_C]$ of size C, an outcome vector $\mathbf{f} = \mathbf{f}(\mathbf{y}) = [f_1, \cdots, f_C]$ is defined.

The RB-pool allocation aims to assign to each cluster as many RBs as possible while assuring proportionality to the cluster resource demand and limiting the inter-cluster interference to an acceptable value. Therefore, the RB-pool allocation problem can be interpreted as a max-min fairness optimization problem [62] consisting in lexicographical maximization of the sorted outcome vector and can be formulated as

$$\hat{\mathbf{Y}}_{\text{opt}} = \arg \operatorname*{lexmax}_{\mathbf{Y} \in (0,1)^{C \times K}} \left\langle \mathbf{f}\left(\mathbf{u}\mathbf{Y}^{\text{T}}\right) \right\rangle \tag{4.2}$$

subject to

$$Y_{c,k} \sum_{c' \in \Omega_{1,C} \setminus \{c\}} Y_{c',k}\, \bar{\iota}_{\text{clust},c,c'} \leq \bar{\psi}_{\text{I}} \;\; \forall (c,k) \in \Omega_{1,C} \times \Omega_{1,K}, \tag{4.2a}$$

where $\mathbf{u}$ is a K-dimensional row vector of ones. The constraint (4.2a) limits the slow CSI-based inter-cluster interference to $\bar{\psi}_{\mathrm{I}}$.

4.2.1.3 Dynamic Intra-Cluster Resource Scheduling

Upon clustering of multicast groups and allocating an RB-pool to each cluster, intra-cluster dynamic resource allocation is performed at the vehicle level, either by a CH V-UE or in a distributed manner within the cluster, acquiring CSI and QSI from the cluster members. Let $\mathcal{F}_{c,n} = \{m \in \mathcal{M}_{\mathrm{clust},c} : Q_{m,n} > 0\}$ be the set of indices of all multicast groups in the c^{th} cluster whose queue is non-empty at time slot n. Since the resource reuse is handled by the RC function during RB-pool assignment and V-UEs belonging to the same cluster are in close proximity, an RB allocation can be restricted to at most one multicast group at a time slot within each cluster to allow for a simpler implementation of a dynamic scheduler at the VEC node. Therefore, for a desired signal transmission from a transmitter V-UE of the m^{th} multicast group with $m \in \mathcal{M}_{\mathrm{clust},c}$ over the V2V link $\mathrm{L}_{d,\tilde{d}}$ at time slot n in the k^{th} RB within a cluster, the worst-case interference power at an intended V-UE receiver $\tilde{d}$ with $\tilde{d} \in \mathcal{D}_{\mathrm{rx},m}$ reads

$$\iota_{\tilde{d},k,n} = \sum_{c' \in \Omega_{1,C} \setminus \{c\}} Y_{c',k} P_{\mathrm{data}} \left(\max_{\substack{\mathrm{L}_{d_{\mathrm{tx},m',\tilde{d}}} \in \mathcal{L}, \\ m' \in \mathcal{M}_{\mathrm{clust},c'}}} H_{\mathrm{L}_{d_{\mathrm{tx},m',\tilde{d}},k,n}} \right). \tag{4.3}$$

Then, considering the worst-case channel quality denoted as $\dot{\gamma}_{m,k,n}(\mathbf{\Pi})$ and by taking into account the worst-case interference $\iota_{\tilde{d},k,n}$ in $\gamma_{\mathrm{L}_{d,\tilde{d}},k,n}(\mathbf{\Pi})$ in (2.3), the intra-cluster resource allocation problem within the c^{th} cluster reads

$$\hat{\mathbf{\Pi}}_{\mathrm{opt},n,c} = \arg \max_{\mathbf{\Pi} \in \Omega_{0,A+1}^{M \times K}} \sum_{m \in \mathcal{F}_{c,n}} u_{\mathrm{del},m} \left(\tau_{m,n}(\mathbf{\Pi}) \right) \tag{4.4}$$

subject to

$$\Pi_{m,k} = 0 \quad \forall\,(m,k) \in \Omega_{1,M} \backslash \mathcal{F}_{c,n} \times \Omega_{1,K} \tag{4.4a}$$

$$\sum_{m \in \mathcal{F}_{c,n}} \mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) \leq 1 \quad \forall k \in \Omega_{1,K} \tag{4.4b}$$

$$Y_{c,k} - \mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) \geq 0 \quad \forall\,(m,k) \in \mathcal{F}_{c,n} \times \Omega_{1,K} \tag{4.4c}$$

$$\sum_{m \in \mathcal{M}_{\mathrm{CAD},x}} \max_{k \in \Omega_{1,K}} \mathbb{1}_{\Omega_{1,A}}(\Pi_{m,k}) \leq 1 \quad \forall x \in \Omega_{1,X} \tag{4.4d}$$

$$\gamma_{\mathrm{thr},m}(\Pi_{m,k}) - \dot{\gamma}_{m,k,n}(\mathbf{\Pi}) \leq 0 \quad \forall (m,k) \in \mathcal{F}_{c,n} \times \Omega_{1,K}. \tag{4.4e}$$

Here, (4.4a) restricts the resource allocation only to the multicast groups residing in the cluster with non-empty queues, while (4.4b) restricts an RB allocation to only at most one multicast group within the cluster. Furthermore, (4.4c) assures that only those RBs being assigned to the RB-pool are allocated, and (4.4d) assures that only at most one multicast group is scheduled in a CAD group at a given time slot to avoid packet loss due to the half-duplex limitation. Finally, (4.4e) assures that the worst-case channel quality is at least the required channel quality so that the reliability requirements of multicast groups are satisfied.

4.2.2 Graph-Theoretic Solution

Due to the combinatorial nature of the problem formulated in Sect. 4.2.1, a graph-based framework is presented for solving the three-stage resource allocation problem with reduced complexity to allow for simpler implementation and enhanced modularity.

Clustering of Multicast Groups:

With constraint (4.1c) that implies all the multicast groups that belong to the same CAD group must always be clustered together, multicast groups clustering can be

viewed as clustering of CAD groups. Then, considering (4.1a), two CAD groups can be clustered together only if all the multicast groups involved can reach each other on the control plane, that is, the x^{th} and x'^{th} CAD groups can potentially reside in the same cluster if

$$Z_{m,m'} \geq \bar{\psi}_{\text{ctrl}} \; \forall (m, m') \in \mathcal{M}_{\text{CAD},x} \times \mathcal{M}_{\text{CAD},x'}. \tag{4.5}$$

Furthermore, for a set of all CAD groups $\mathcal{X} \subseteq \Omega_{1,X}$ that are part of the same cluster, the constraint (4.5) must always hold $\forall (x, x') \in \mathcal{X}^2$. Representing each CAD group as a vertex of a graph and defining two vertices to be adjacent, i.e., having an edge, if (4.5) holds for the corresponding CAD groups, a cluster of CAD groups can be interpreted as a sub-graph in which all the vertices are adjacent to each other. In graph theory, such a sub-graph is called a clique [63]. Let $\mathrm{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with vertex set $\mathcal{V}$ and edge set $\mathcal{E}$. Here, the vertex $\mathrm{V}_i \in \mathcal{V}$ indicates the i^{th} CAD group, and $\mathrm{E}_{\mathrm{V}_i,\mathrm{V}_j} \in \mathcal{E}$ an edge between vertices V_i and V_j if (4.5) holds for the pair of the i^{th} and j^{th} CAD groups. Furthermore, in order to define affinity among multicast groups across CAD groups, an edge attribute is introduced as

$$W_{\mathrm{E}}\left(\mathrm{E}_{\mathrm{V}_i,\mathrm{V}_j}\right) \; = \sum_{m \in \mathcal{M}_{\text{CAD},i}} \sum_{m` \in \mathcal{M}_{\text{CAD},j}} W_{m,m`}.$$

Then, the solution to the multicast groups clustering problem in (4.1) can be translated to finding disjoint sets of vertices inducing cliques (clusters) that partitions G as illustrated in Fig. 4.7 such that the sum of edge weights in every clique is maximized. In other words, defining the clique weight as the sum of the weights of the edges within the clique, the clustering problem is to partition the graph into cliques (clusters) such that the sum of weights of the cliques is maximized. Such a graph partitioning problem is known as a CPP [63], [64], where [63] considers the minimization of the sum of the clique weights as the clustering objective. Since the

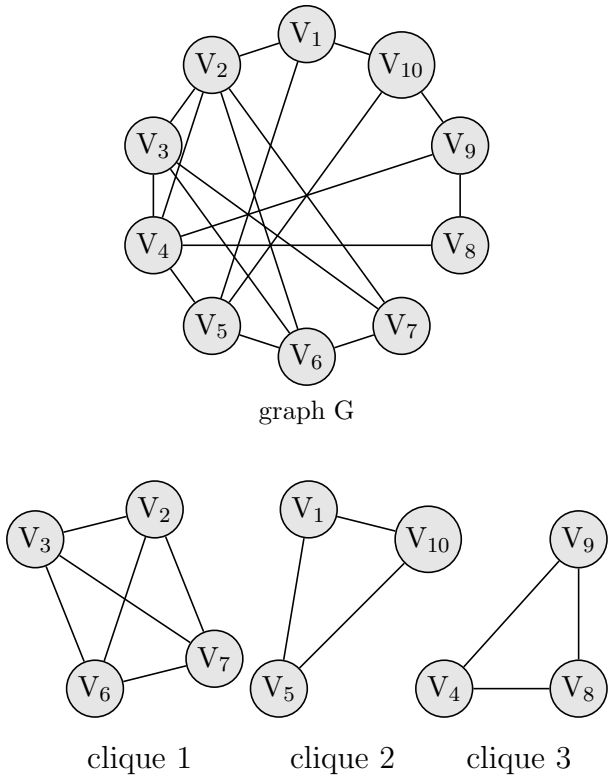


Figure 4.7: Illustration of partitioning inter-CAD group control plane reachability graph G with 10 CAD groups into three cliques.

CPP is known to be NP-hard [63], the heuristic **Algorithm** 4 is proposed, where all vertices of graph G are partitioned into a set of clusters as cliques with a high clique value. The algorithm initializes each CAD group to be a distinct cluster and merges clusters progressively based on the affinity among the CAD groups. In this regard, for a given number of clusters (cliques) C, an averaged inter-cluster affinity weight is defined for the c^{th} and c'^{th} cluster with $(c, c') \in \Omega_{1,C}^2$ as

$$W_{\text{clust},c,c'} = \left(|\mathcal{V}_c|\,|\mathcal{V}_{c'}|\right)^{-1} \sum_{(\text{V}_i,\text{V}_j)\in\mathcal{V}_c\times\mathcal{V}_{c'}} W_{\text{E}}\left(\text{E}_{\text{V}_i,\text{V}_j}\right) \left(\prod_{(\text{V}_i,\text{V}_j)\in\mathcal{V}_c\times\mathcal{V}_{c'}} \mathbb{1}_{\mathcal{E}}(\text{E}_{\text{V}_i,\text{V}_j}) \right),$$

where $\mathcal{V}_c \subset \mathcal{V}$ is the set of all vertices in the c^{th} clique. The algorithm avoids to merge two clusters if the involved multicast groups cannot coordinate jointly and the aforementioned expression results in $W_{\text{clust},c,c'} = 0$.

Algorithm 4: Clustering of Multicast Groups

Input : G

Output: Sets of vertices $\mathcal{V}_{\tilde{c}}, \forall \tilde{c} \in \tilde{\mathcal{C}} \subseteq \Omega_{1,|\mathcal{V}|}$, which form distinct cliques

Initialization: $\tilde{\mathcal{C}} \leftarrow \Omega_{1,|\mathcal{V}|}, \mathcal{V}_{\tilde{c}} \leftarrow \{\mathrm{V}_{\tilde{c}}\} \ \forall \tilde{c} \in \Omega_{1,|\mathcal{V}|}$;

do

  $(\hat{c}, \hat{c}') \leftarrow \arg \max_{(c,c') \in \tilde{\mathcal{C}}^2} W_{\mathrm{clust},c,c'}$;

  if $W_{\mathrm{clust},\hat{c},\hat{c}'} > 0$ **then**

    $\mathcal{V}_{\hat{c}} \leftarrow \mathcal{V}_{\hat{c}} \cup \mathcal{V}_{\hat{c}'}$;

    $\tilde{\mathcal{C}} \leftarrow \tilde{\mathcal{C}} \setminus \{\hat{c}'\}$;

  else

    break;

  end if

while $\left|\tilde{\mathcal{C}}\right| > 1$;

Inter-Cluster RB-Pool Allocation:

Each cluster can be viewed as a vertex of an undirected complete resource-conflict graph $\mathrm{G}^{\mathrm{\backslash}} = (\mathcal{V}^{\mathrm{\backslash}}, \mathcal{E}^{\mathrm{\backslash}})$, where each edge $\mathrm{E}^{\mathrm{\backslash}}_{\mathrm{V}^{\mathrm{\backslash}}_{i^{\mathrm{\backslash}}}, \mathrm{V}^{\mathrm{\backslash}}_{j^{\mathrm{\backslash}}}} \in \mathcal{E}^{\mathrm{\backslash}}$ between a pair of cluster vertices $\mathrm{V}^{\mathrm{\backslash}}_{i^{\mathrm{\backslash}}}$ and $\mathrm{V}^{\mathrm{\backslash}}_{j^{\mathrm{\backslash}}}$ with $\left(\mathrm{V}^{\mathrm{\backslash}}_{i^{\mathrm{\backslash}}}, \mathrm{V}^{\mathrm{\backslash}}_{j^{\mathrm{\backslash}}}\right) \in {\mathcal{V}^{\mathrm{\backslash}}}^2$ can be weighted to identify the inter-cluster worst case interference. The vertices can be weighted with a corresponding cluster weight to account for the cluster resource demand. In this regard, three attributes are introduced on the graph as follows:

- a demand attribute $W_{\mathrm{V}}(\mathrm{V}^{\mathrm{\backslash}}_{i^{\mathrm{\backslash}}}) = \Gamma_{i^{\mathrm{\backslash}}}$ as the aggregated resource demand of the $i^{\mathrm{\backslash th}}$ cluster
- an RB-pool attribute $\mathcal{K}_{\mathrm{V}^{\mathrm{\backslash}}_{i^{\mathrm{\backslash}}}} \in \Omega_{1,K}$ being the set of RBs assigned to the $i^{\mathrm{\backslash th}}$ cluster
- an interference attribute $I\left(\mathrm{E}^{\mathrm{\backslash}}_{\mathrm{V}^{\mathrm{\backslash}}_{i^{\mathrm{\backslash}}}, \mathrm{V}^{\mathrm{\backslash}}_{j^{\mathrm{\backslash}}}}\right) = \bar{\iota}_{\mathrm{clust},i^{\mathrm{\backslash}},j^{\mathrm{\backslash}}}$ representing the inter-multicast group worst-case interference between the $i^{\mathrm{\backslash th}}$ and $j^{\mathrm{\backslash th}}$ clusters.

The inter-cluster RB-pool allocation in Sect. 4.2.1.2 can be translated to the task of assigning an optimal number of RBs in the RB-pool attribute of vertices on the resource-conflict graph such that, at no vertex, the sum of interference attributes over the edges with vertices sharing the RB exceeds $\bar{\psi}_{\mathrm{I}}$. In addition, since the RB-pool allocation assures max-min fairness, **Algorithm** 5 provides a heuristic approach to allocate an RB-pool to each cluster vertex proportionately to

Algorithm 5: Inter-Cluster RB-pool Allocation

Input : $\mathrm{G}^{\backprime}$

Output: $\mathcal{K}_{\mathrm{V}^{\backprime}}\ \forall \mathrm{V}^{\backprime} \in \mathcal{V}^{\backprime}$

Initialization:

$\mathcal{K}_{\mathrm{V}^{\backprime}} \leftarrow \emptyset\ \forall \mathrm{V}^{\backprime} \in \mathcal{V}^{\backprime}, \mathcal{V}^{\backprime}(k) \leftarrow \emptyset\ \forall k \in \Omega_{1,K}, \check{\mathcal{V}}^{\backprime} \leftarrow \emptyset,$

$\check{W}_{\mathrm{V}}(\mathrm{V}^{\backprime}) \leftarrow 1/W_{\mathrm{V}}(\mathrm{V}^{\backprime})\ \forall \mathrm{V}^{\backprime} \in \mathcal{V}^{\backprime};$

for $k \leftarrow 1 : K$ **do**

  do

    $\hat{\mathrm{V}}^{\backprime} \leftarrow \underset{\mathrm{V}^{\backprime} \in \mathcal{V}^{\backprime} \setminus \check{\mathcal{V}}^{\backprime}}{\arg\ \min}\, \check{W}_{\mathrm{V}}(\mathrm{V}^{\backprime});$

    $\tilde{\mathcal{V}}^{\backprime}(k) \leftarrow \mathcal{V}^{\backprime}(k) \cup \{\hat{\mathrm{V}}^{\backprime}\};$

    if $\mathcal{V}^{\backprime}(k) = \emptyset$ **or** $\sum_{\mathrm{V}^{\backprime} \in \tilde{\mathcal{V}}^{\backprime}(k) \setminus \{\hat{\mathrm{V}}^{\backprime}\}} I\left(\mathrm{E}^{\backprime}_{\mathrm{V}^{\backprime}, \tilde{\mathrm{V}}^{\backprime}}\right) \leq \bar{\psi}_{\mathrm{I}}, \forall \tilde{\mathrm{V}}^{\backprime} \in \mathcal{V}^{\backprime}(k)$ **then**

      $\mathcal{K}_{\hat{\mathrm{V}}^{\backprime}} \leftarrow \mathcal{K}_{\hat{\mathrm{V}}^{\backprime}} \cup \{k\};$

      $\mathcal{V}^{\backprime}(k) \leftarrow \mathcal{V}^{\backprime}(k) \cup \{\hat{\mathrm{V}}^{\backprime}\};$

    end if

    $\check{W}_{\mathrm{V}}\left(\hat{\mathrm{V}}^{\backprime}\right) \leftarrow \left|\mathcal{K}_{\hat{\mathrm{V}}^{\backprime}}\right| / W_{\mathrm{V}}(\hat{\mathrm{V}}^{\backprime});$

    $\check{\mathcal{V}}^{\backprime} \leftarrow \check{\mathcal{V}}^{\backprime} \cup \{\hat{\mathrm{V}}^{\backprime}\};$

  while $\left|\check{\mathcal{V}}^{\backprime}\right| < |\mathcal{V}^{\backprime}|;$

  $\check{\mathcal{V}}^{\backprime} \leftarrow \emptyset;$

end for

the corresponding demand attribute in a progressive filling manner [62].

Intra-Cluster Resource Allocation:

Upon clustering multicast groups and allocating an RB-pool to each cluster, intra-

cluster dynamic resource allocation is performed by solving the constrained optimization problem in (4.4). The clustering of multicast groups will lead to a lower number of CAD groups per cluster, and at most one multicast group in each CAD group within the cluster can be scheduled due to the limitations imposed by the half-duplex constraint. In addition, orthogonal RBs are allocated between adjacent clusters in the RB-pool allocation and hence no coordination is needed across clusters. Furthermore, due to the close proximity of CAD groups within a cluster, the intra-cluster allocation of an RB is restricted to at most one multicast group (i.e., no RB reuse within a cluster). Therefore, by nominating a multicast group from each CAD group for resource allocation based on the utility at a given time slot (as shown in **Algorithm** 1), the intra-cluster resource allocation problem can be solved per cluster using a simple iterative heuristic algorithm as shown in **Algorithm** 6 or a conventional exhaustive-search algorithm [65] with a substantially reduced search space. Here, dynamic radio resource allocation can be performed by any of the vehicles within the in order to as a CH scheduler using CSI and QSI from the cluster members.

In this work, a CM function is assumed to have randomly selected a V-UE as the CH for each cluster upon clustering V-UEs. Each CH is provided with information from the CM which includes the class identity, multicast groups (including the involved V-UEs) belonging to the cluster with associated QoS-dependent utility functions and the CAD group information, and the potential transmitter V-UEs in each of other clusters. Furthermore, a RC function determines a CH with RB-pool assignment information of all clusters (or at least the clusters which reuse the same RBs).

A simple heuristic approach for RRS at the c^{th} cluster by a CH is shown in **Algorithm** 6, which makes use of **Algorithm** 1 as described in Sect. 3.2.2 to reduce the search space by nominating a multicast group from each CAD group within

Algorithm 6: RRS Algorithm for c^{th} Cluster

Input : full CSI and QSI with $\check{\mathcal{F}}_{c,n}$
Output: $\mathbf{\Pi}$

Initialization: $\mathbf{\Pi} \leftarrow \mathbf{0}_{M\times K}$;
foreach $k \in \mathcal{K}_c$ **do**
  $\mathcal{M} \leftarrow \check{\mathcal{F}}_{c,n}$;
  do
    $\acute{m} \leftarrow \arg\max_{\tilde{m}\in\mathcal{M}} -u'_{\text{del},\tilde{m}}\left(\tau_{\tilde{m},n}(\mathbf{\Pi})\right)$;
    for $a = A : 1$ **do**
      $\tilde{\mathbf{\Pi}} \leftarrow \mathbf{\Pi}$;
      $\tilde{\Pi}_{\tilde{m},k} \leftarrow a$;
      if $\gamma_{\text{thr},\acute{m}}(\tilde{\Pi}_{\tilde{m},k}) - \dot{\gamma}_{\acute{m},k,n}(\tilde{\mathbf{\Pi}}) \leq 0$ **then**
        $\Pi_{\acute{m},k} \leftarrow \tilde{\Pi}_{\tilde{m},k}$;
        break;
      end if
    end for
    $\mathcal{M} \leftarrow \mathcal{M} \setminus \{\acute{m}\}$;
  while $\Pi_{\acute{m},k} = 0$ **and** $|\mathcal{M}| > 0$;
end foreach

a cluster and therefore assuring that (4.4d) is satisfied. Let $\check{\mathcal{F}}_{c,n}$ be the set of nominated multicast groups for resource allocation in the c^{th} cluster whose queue is non-empty at time slot n. Iterating over RBs, the algorithm assigns RBs from the corresponding RB-pool incrementally to the multicast groups in $\check{\mathcal{F}}_{c,n}$ by selecting a multicast group for allocation in each RB using a utility-based greedy criterion. Here, similarly to **Algorithm** 2, the algorithm considers the negative slope (as the packet delay is a downward criterion) of the utility function $-u'_{\text{del},m}\left(\tau_{m,n}(\mathbf{\Pi})\right)$ for the HOL delay at the end of the current time slot $\tau_{m,n}(\mathbf{\Pi})$ for the given allocation $\mathbf{\Pi}$ as a greedy criterion since the objective of (4.4) is to maximize the aggregated utility of all multicast groups within the cluster. Moreover, for the chosen multicast

group, the algorithm also allocates the best MCS (i.e., highest-order MCS) that can be supported in the RB considering worst-case SINR performance and reliability requirements.

4.2.3 Advantages and Challenges

The edge cloud-enabled three-stage resource allocation framework splits the resource allocation into long-term RRM and dynamic RRS tasks. The long-term RRM is performed at the central cloud server whereas the dynamic RRS is performed at a distributed VEC node at vehicle level. Finding solutions to the proposed three-stage resource allocation problems requires V-UEs to report slow CSI to the cloud server and full CSI and QSI to neighbouring V-UEs (e.g., to a CH). Being an edge cloud-based scheme, it inherits the benefits of cloud computing, such as its ability to be operator-independent, and allows for resource control by third-parties, e.g., automotive organizations.

The proposed three-stage semi-centralized scheme provides dynamic RRS at vehicular level by placing the scheduler at neighboring or co-located VEC nodes. Hence, as illustrated in Fig. 4.8 the dynamic scheduling delay is significantly reduced due to the close proximity of the scheduler that offers a low V2V link delay (or V2V delay for short). This allows the scheme to support low latency CAD applications such as CCA. Also, because of the QoS-dependent utility-based RRS, the scheme can cater to diverse QoS requirements.

Similar to the case of slow CSI-based centralized RRS described in Sect. 3.3, because the V-UEs need to report only the slow CSI in a long-term basis at an interval of $T_{\mathrm{c}} \in \mathbb{N}$ time slots as opposed to reporting the full CSI at every time slot, the signaling overhead over V2N is significantly reduced. Also, since no QSI is reported to the sidelink RRM entities in the central cloud server, the V2N signaling overhead is reduced further. Furthermore, by reusing RBs across clusters, the resource uti-

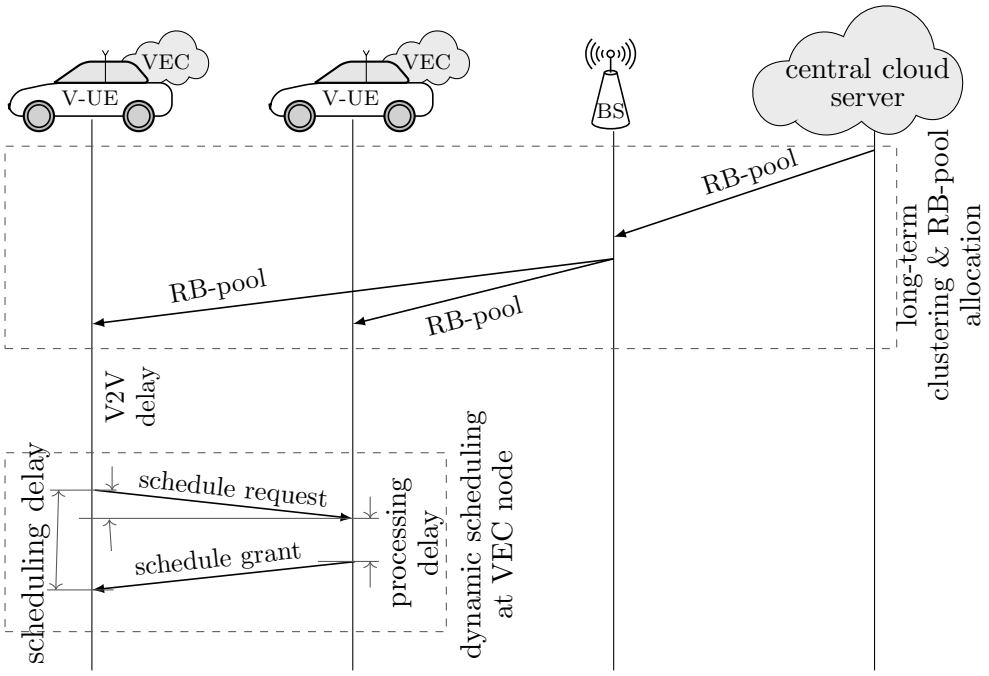


Figure 4.8: Illustration of scheduling delay in three-stage resource allocation for a cluster of V-UEs

lization is increased. At the same time, since the dynamic RRS is performed at the vehicle level, the reliability performance on V2V links is not impacted by the high V2N delays. However, since the intra-cluster dynamic scheduler collects full CSI and QSI over V2V links, the scheme introduces significant signaling overhead over V2V links.

Chapter 5

Performance Evaluations for Platooning in a 5G Set-Up

In order to evaluate the performance of the proposed sidelink radio resource allocation schemes for CAD V2V communications, a realistic dynamic vehicular topology for a platooning use case is simulated considering two high and low LoA in a 5G-based simulation set-up. Here, the vehicles travel in groups as platoons and within each CAD group (or platoon), the trailing vehicles follow the lead vehicle as illustrated in Fig. 1.1. Each CAD group is considered to be running with either high or low LoA, where high and low LoA indicate self-driven and semi-autonomously driven vehicles, respectively. In the following, Sect. 5.1 describes the simulation set-up and the performance metrics under evaluation. Simulation results are discussed for centralized full CSI and slow CSI-based dynamic RRS schemes in Sect. 5.2 and for the three-stage resource allocation in Sect. 5.3.

5.1 5G-based Simulation Set-Up

The microscopic traffic simulator SUMO [66] is used to generate a realistic vehicular mobility. Here, vehicles are deployed as platoons uniformly with an average velocity of 25 m/s on a highway segment described for 5G evaluations in [67] as illustrated in Fig. 5.1. The road segment is considered to be 5000 m in length with 3 lanes

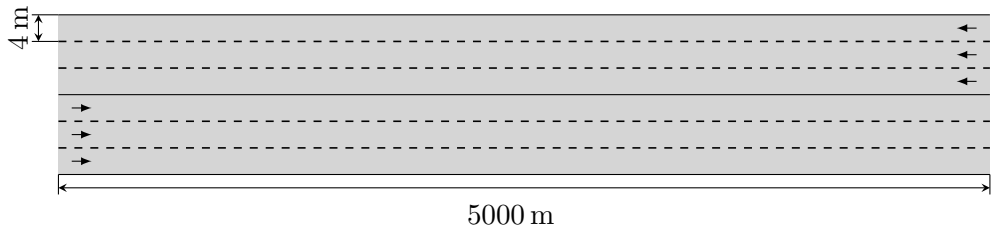


Figure 5.1: Highway road configuration.

in each direction (6 lanes in total) with a lane width of 4 m. For convenience of simulation, the vehicles are made to perform a U-turn upon reaching the end of the road segment so that they do not leave the simulated road area.

The scenario considers 94 vehicles where each vehicle contains a V-UE to perform V2V communication within the CAD group to implement the platooning operation. Furthermore, the vehicles are assumed to have already formed 27 platoons (or CAD groups) and exchange CAD messages either periodically or non-periodically within the respective platoon using V-UEs over V2V links. That is, within each CAD group, the V-UEs perform multicast transmissions (i.e., groupcasts) to all the other member V-UEs. Hence, there are 94 multicast groups since there are 94 vehicles. Among all CAD groups, 13 groups (i.e., about half of the groups) consist of 4 member V-UEs and 14 groups consist of 3 member V-UEs. Moreover, a sub-set of 9 (i.e., a third of the platoons) CAD groups are randomly selected as high LoA platoons, whereas the rest is assumed to operate with low LoA.

5G considers 90% and 99.99% as reliability requirements and 25 ms and 10 ms as maximum end-to-end latencies for low and high LoA platoons, respectively [12]. In the evaluation, an inelastic (strict) requirement on the reliability ρ for each CAD V2V multicast group transmission is assumed where a reliability below ρ_{req} is considered unacceptable. Here, ρ_{req} is assumed to be 90% and 99.99% for low and high LoA platoons, respectively. Furthermore, instead of a strict delay requirement, a partially elastic requirement on the delay τ in terms of sigmoid function is assumed

to reflect the performance tolerance at the cost of platoon operation efficiency as described in Sect. 3.1.3. With a steepness parameter ϵ and a center ζ of the sigmoid function, an exemplary utility is assumed to be $1 - (1 + \exp(\epsilon(\zeta - \tau)))^{-1}$ if $\rho \geq \rho_{\text{req}}$ as illustrated in Fig. 5.2 and 0 otherwise. Here, ϵ is set to 0.85 and 0.65 for

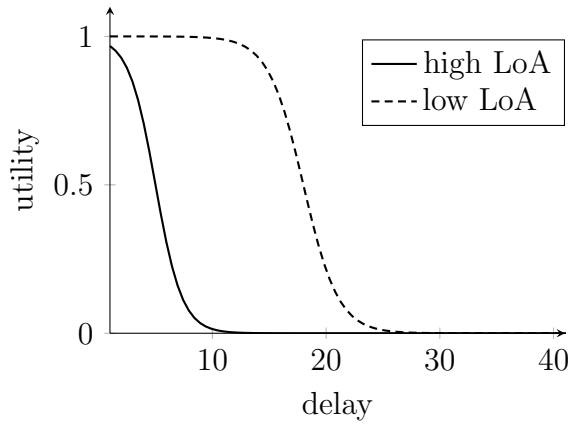


Figure 5.2: Utility function capturing example QoS requirement of high and low LoA platooning.

high and low LoA, resp., where the higher steepness for high LoA reflects the higher sensitivity to change in delay performance during platooning due to the higher criticality of the platoon operation under high LoA. Moreover, ζ is set to 5 ms and 18 ms for high and low LoA, respectively. With this setting, utility of a multicast group drops closer to 0 when the delay approaches 10 ms and 25 ms for high and low LoA, respectively, to reflect the end-to-end delay requirements considered in 5G.

For a packet size of 400 bytes and 300 bytes, a transmitter V-UE of each multicast group is considered to transmit 100 and 30 packets per second within the CAD group in case of high and low LoA, resp., in line with 5G specifications [12].

With a dedicated bandwidth of 3 MHz for CAD in the 5.9 GHz band and a time slot duration T_{sl} of 1 ms, an RB is considered to span 12 subcarriers with a subcarrier spacing (SCS) of 15 kHz (hence, 15 RBs available in 3 MHz bandwidth) and consists of 150 REs for payload. Furthermore, the sidelink transmission powers for data P_{data} and control P_{ctrl} are considered to be 5 dBm and 15 dBm, respectively. The V2V

sidelink channel model described by 3GPP for 5G in [67] is used. Aligning with the 5G implementation, the evaluation employs low-density parity-check (LDPC) coding with MCS information presented in Table 5.1, where the SINR thresholds

Table 5.1: MCS information.

	MCS-1	MCS-2	MCS-3	MCS-4	MCS-5	MCS-6
$S(a)$	2	2	4	4	6	8
$R(a)$	0.468	0.746	0.598	0.950	0.833	0.889
$\gamma_{\text{thr}}(a)$ [dB] ($\rho_{\text{req}} = 90.00\%$)	0.1	3.0	7.0	12.1	16.0	23.7
$\gamma_{\text{thr}}(a)$ [dB] ($\rho_{\text{req}} = 99.99\%$)	0.8	3.7	7.8	12.9	17.6	24.6

for different MCSs and reliability requirements are identified based on the BLER performance.

Moreover, the delay over the V2N link T_{V2N} is assumed to be fixed to 50 ms. Each V-UE is considered to report slow CSI to the central cloud server at an interval T_{c} of 100 ms when applicable. Table 5.2 summarizes the default simulation parameters.

The metrics under evaluation comprise:

a) the cumulative distribution function (CDF) of the SINR on V2V links which captures the per-link distribution of the channel quality.

b) the packet delivery ratio (PDR) across multicast groups to capture the reliability performance, wherein PDR is the ratio of the number of successfully received sidelink multicast group transmissions (i.e., the number of successful transport block (TB) transmissions) and the total number of sidelink multicast group transmissions. Here, a multicast group transmission is considered to be successful if the channel quality for the corresponding transmission of the multicast group meets the reliability requirement in terms of the channel quality requirement (3.3) of the group for the used MCS.

Table 5.2: Simulation parameters.

Parameters		**Value**
carrier frequency		5.9 GHz
bandwidth		3 MHz
no. of RBs	K	15
no. of slots	N	10000
slot duration	T_{sl}	1 ms
no. of REs in a RB for payload	K_{RE}	150
road topology		Highway as shown in Fig. 5.1
deployment	D	94 V-UEs
	X	27 platoons uniformly distributed on highway; 13 groups with 4 members; 14 groups with 3 members; 33% of groups are high LoA
	M	94 multicast groups
average vehicle velocity		25 m/s
no. of MCSs	A	6
packet generation rate	$T_{\mathrm{pkt},m}$	100 packets/s for high LoA; 30 packets/s for low LoA
packet size	B_m	400 bytes for high LoA; 300 bytes for low LoA
channel model		As defined by 3GPP[67]
transmission power	P_{data}	5 dBm (data)
	P_{ctrl}	15 dBm (control)
V2N delay	T_{V2N}	50 ms
slow CSI reporting interval	T_{c}	100 ms
noise power	σ^2	-170 dBm

c) the packet delay with adaptive modulation and coding captured at the low (10th), median (50th) and high (90th) percentile of the delay CDF to reflect the delay distribution. Here, each multicast group transmission is assumed to be received at the intended recipients successfully.

The following schemes are simulated for the performance evaluation.

- Full CSI-based centralized RRS: A full CSI-based (i.e., instantaneous CSI-based) dynamic centralized RRS mechanism as described in Sect. 3.2 where the scheduler is placed at the central cloud server. V-UEs report CSI and QSI at every time slot to the scheduler over V2N links. The scheme is susceptible to the challenges posed by large V2N delays as it risks the report being outdated.

- Slow CSI-based centralized RRS: A slow CSI-based dynamic centralized RRS mechanism as described in Sect. 3.3 where the scheduler is placed at the central cloud server. Unlike the full CSI-based scheme, V-UEs report only slow CSI over V2N links to the scheduler every T_{c} time slots to be robust against the V2N delay. The slow CSI-based RRS provides an extra SINR margin to account for the potential deviation of the SINR performance due to the influence of fast fading which is not available in the slow CSI measurements, however, at the cost of a resource utilization (i.e., a lower RB reuse). Here, the extra SINR margin provided for multicast groups corresponding to low (for 90% reliability) and high LoA (for 99.99% reliability) is 9.77 dB and 39.99 dB, resp., as per (3.8). The scheme maintains a virtual queue to track QSI for each multicast group and hence the V-UEs do not need to report QSI which anyway would have been outdated in case of a large V2N delay. Here, NPDT is expected to pose a challenge since the QSI cannot be accurately tracked at the scheduler.

- Three-stage semi-centralized RRS: A semi-centralized three-stage resource allocation as described in Sect. 4.2. The long-term RRM tasks, namely the clustering of V-UEs and the inter-cluster RB-pool allocation, are performed at the central cloud server every T_{c} time slots based on slow CSI reported by V-UEs over V2N links. Whereas, a full CSI-based dynamic RRS is performed at the vehicle level by a CH V-UE, and member V-UEs of each cluster report

CSI and QSI continuously to the respective CH over V2V links. Because of the close proximity of the scheduler and hence a reduced scheduling delay, the scheme is expected to provide low latency performance.

- Ideal RRS: The full CSI-based centralized RRS scheme with perfect knowledge of instantaneous CSI and QSI available at the scheduler is considered to be an ideal case. Hence, the ideal RRS scheme assumes zero delay on V2N links (i.e., $T_{\mathrm{V2N}} = 0$).

- Orthogonal RRS: An LTE-V Mode-3 [68] semi-persistent scheduling (SPS)-based network-controlled conservative fully orthogonal RB allocation scheme with a round-robin scheduler as a benchmark (BM). In this case, different RBs are assigned to each multicast group so that no interference is expected.

5.2 Centralized Dynamic Resource Scheduling at Cloud Server

At first, the full CSI-based centralized dynamic RRS scheme is simulated with no control plane delay on V2N links to evaluate the ability of the utility-based centralized scheduler to flexibly adapt with the available bandwidth of different sizes in order to cater to CAD applications of different delay requirements on V2V links. In this regard, Fig. 5.3 shows a packet delay performance comparison for different bandwidth sizes (i.e., different numbers of RBs available for allocation) assuming a successful reception, where each multicast group is considered to perform PDT transmissions. The packet delay evaluations show that the scheme is able support low latency communication for high LoA even when the bandwidth size is significantly reduced by penalizing delay tolerant low LoA multicast groups although the packet size and generation rate for high LoA are 1.33 and 3.33 times, resp., higher

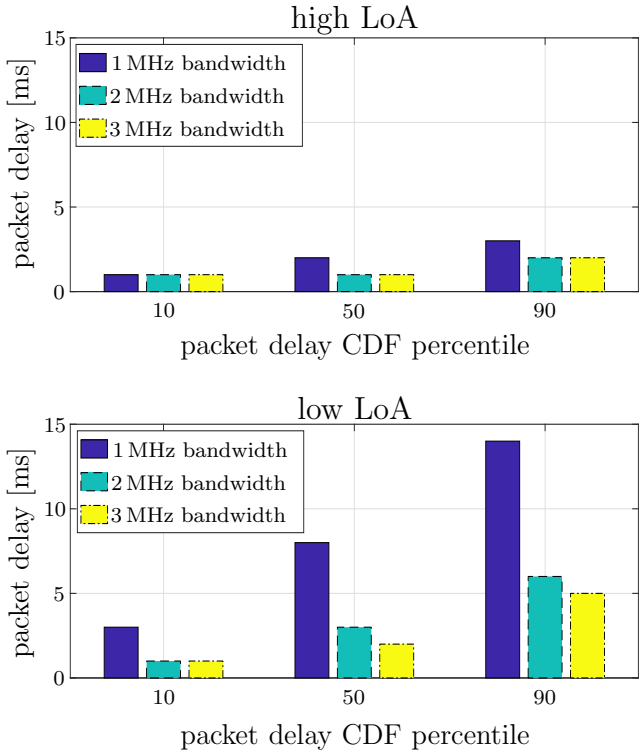


Figure 5.3: Packet delay performance of full CSI-based centralized dynamic RRS for different bandwidth sizes considering no V2N delay and PDT.

than that of low LoA. In particular, at the 90th percentile of the packet delay CDF, even for a bandwidth as low as 1 MHz (i.e., for 5 RBs), the scheme shows desirable 3 ms packet delay performance (high utility of 0.85) for high LoA, while a 15 ms packet delay is observed for the delay-tolerant low LoA which still provides very high utility of 0.88. Also, at the 50th percentile, in spite of only 1 MHz bandwidth, the packet delay performance for high and low LoA is observed to be highly desirable 2 ms (0.92 utility) and 8 ms (0.99 utility), respectively. This shows that the QoS-dependent utility-based scheduler can flexibly support CAD applications even when the resource availability is limited.

5.2.1 Sidelink Reliability for Large V2N Delay

While Fig. 5.3 considers an ideal case of no control plane delay over V2N links, the real-world delay can be significantly higher in the order of tens or hundreds of

milliseconds (e.g., 50 ms) which may potentially impact the QoS performance on V2V links. Fig. 5.4 shows the CDF of the SINR on V2V links involved in multicast group PDT transmissions for full CSI-based RRS schemes for V2N delays of 0 ms and 50 ms. It can be observed that the SINR performance of the full CSI-based scheme

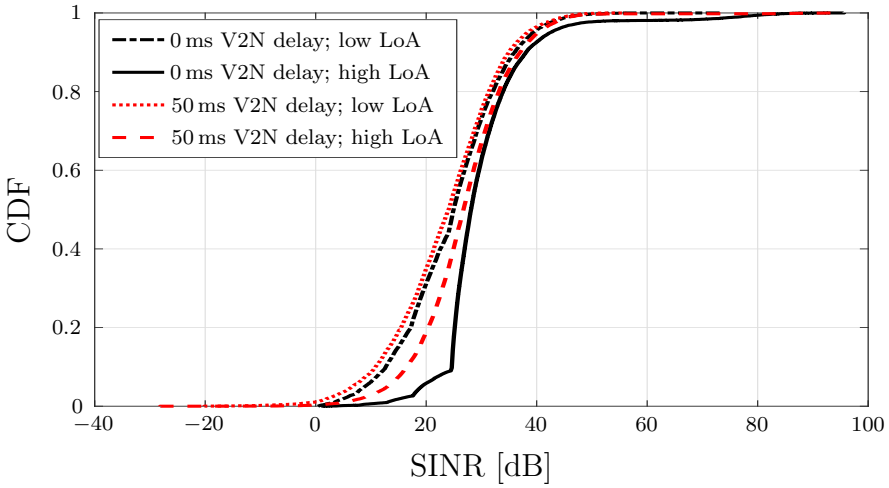


Figure 5.4: SINR performance for full CSI-based centralized scheme for different V2N delays.

is degraded when the V2N delay is increased from 0 ms (i.e., ideal case) to 50 ms (i.e., round-trip delay of 100 ms). In particular, at the 10th percentile of the SINR CDF, a 50 ms V2N delay causes a 1.9 dB and 8.1 dB SINR degradation for low LoA and high LoA, respectively. Also, about 1.1% and 0.4% of the links show an SINR of less than the minimum required SINR of 0.1 dB and 0.8 dB corresponding to the lowest order MCSs for LoA and high LoA, respectively. The SINR degradation is due to the reported CSI being expired during RRS and hence the resource allocation is based on the invalid CSI which does not reflect the actual CSI at the V-UE level at the given time slot. That is, the channel state during the channel access is being different from the reported CSI mainly due to change in fast fading components over the span of 100 ms (considering both a reporting CSI and a receiving schedule grant over the V2N link) which results in increased interference.

Note that Fig. 5.4 shows the actual SINR performance on V2V links upon using

the allocated resources. However, the MCS selection at the scheduler depends on the perceived SINR based on the collected CSI measurements. Hence, the reliability is impacted when the actual SINR is less than the perceived SINR, and therefore the reliability performance in terms of PDR is evaluated. Fig. 5.5 shows the sidelink PDR performance of multicast PDT transmissions associated with both high and low LoA for the full CSI-based centralized RRS schemes for V2N delays of 0 ms and 50 ms. The PDR is observed to be 100% when the V2N delay is assumed to be 0 ms

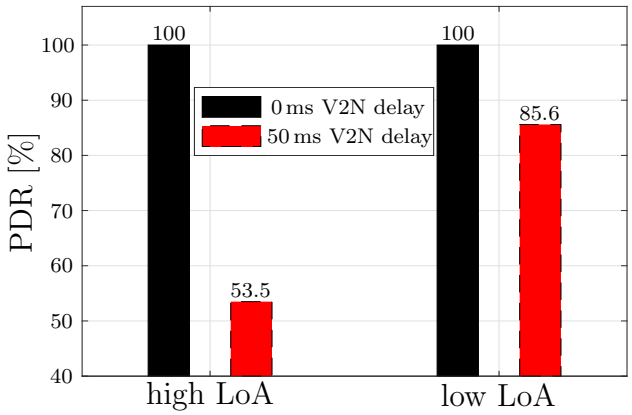


Figure 5.5: PDR performance of full CSI-based centralized RRS scheme for two different V2N control plane delays.

(ideal case) for both LoA, whereas it is about 53% and 86% for high and low LoA, resp., when the V2N link incurs 50 ms (i.e., a round-trip delay of 100 ms). In case of a 0 ms V2N delay (ideal case), the perceived SINR at the scheduler is identical to the actual SINR since the scheduler has perfect knowledge of the CSI. Therefore, it results in 100% PDR. However, in case of a V2N delay of 50 ms, the SINR perceived at the scheduler is based on the expired CSI measurements and hence it differs from the actual SINR. Consequently, the PDR performance is degraded when the actual SINR at the receiving V-UEs is below the minimum required SINR for the MCS that is allocated based on the incorrectly perceived SINR. The PDR performance degradation is significantly higher for high LoA because of the usage of higher-order MCSs that require a high SINR on V2V links aiming to achieve the required

PDR of least 99.99%. Therefore, a drop in the SINR performance significantly impacts the case of high LoA. It is observed that the scheme does not meet the PDR requirements for both high and low LoA when the V2N delay equals 50 ms. To this end, a slow CSI-based centralized RRS can be used to improve the SINR performance as shown in the following. Note that since the type of data traffic (i.e., PDT or NPDT) does not influence the SINR performance, the PDR is identical for PDT and NPDT and hence it is not discussed here.

Fig. 5.6 shows the CDF of the SINR on V2V links involved in multicast group PDT transmissions for the slow CSI-based scheme against the full CSI-based and orthogonal centralized RRS schemes. An orthogonal RRS scheme is considered as a

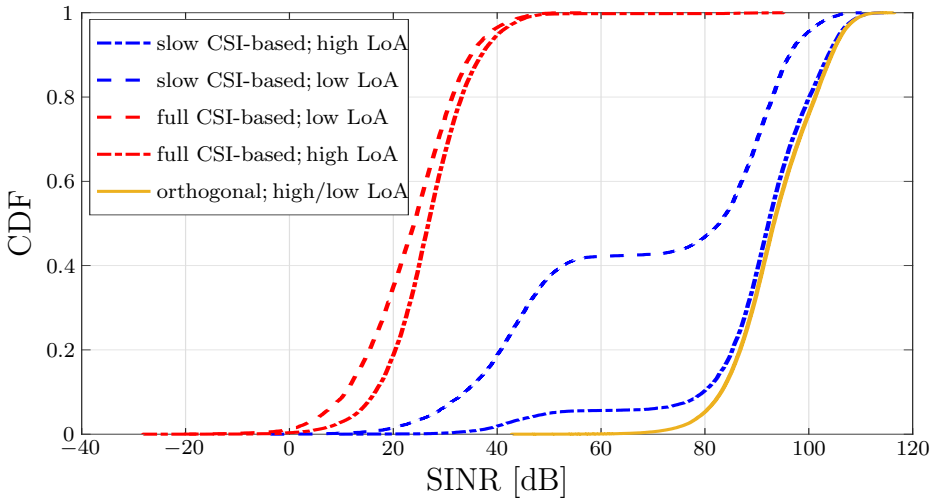


Figure 5.6: SINR performance for different centralized schemes.

BM since it provides the SINR distribution under no RB reuse and hence it naturally outperforms the proposed schemes. Here, the SINR distribution is observed to be the same for both high and low LoA since no interference is expected in either case. The slow CSI-based scheme is observed to offer significantly higher SINR against the full CSI-based scheme which suffers from SINR degradation due to expired CSI as discussed above. In particular, even at the low (10th) percentile of the SINR CDF, the slow CSI-based scheme provides 23 dB and 63 dB gain for low and high LoA,

resp., against the full CSI-based scheme. Unlike the full CSI-based scheme, V-UEs report only slow CSI over V2N links to the scheduler every T_c time slots in order to be robust against the V2N delay, and it provides an extra SINR margin to account for the potential deviation of the actual SINR due to the influence of fast fading that is not available in the slow CSI measurements. Here, the extra SINR margin provided for multicast groups corresponding to low (for 90% reliability) and high LoA (for 99.99% reliability) is 9.77 dB and 39.99 dB, resp., as per (3.8). This, however, comes at the cost of resource utilization (i.e., lower RB reuse). Consequently, even at the 10th percentile, the SINR performance for high LoA already approaches that of the orthogonal scheme since an SINR margin of 39.99 dB significantly restricts the RB reuse. Furthermore, the SINR margin almost inhibits the RB reuse at the center of the road segment and allows for a higher RB reuse among multicast groups as they get away from the center. This results in different SINR regions as observed, in particular, near the median of the CDF for low LoA.

Fig. 5.7 shows the PDR performance of multicast group PDT transmissions associated with both high and low LoA for the centralized RRS schemes. The orthogonal

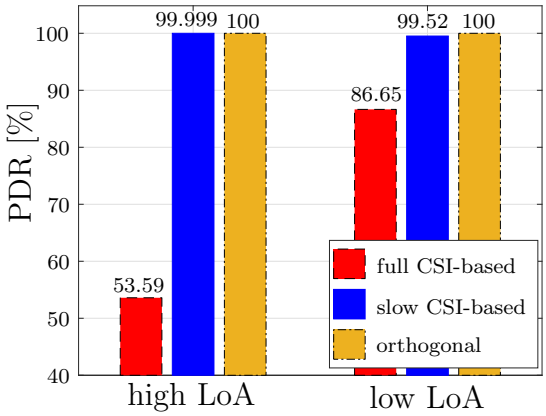


Figure 5.7: PDR performance for different centralized schemes.

scheme outperforms the proposed schemes since it assigns orthogonal RBs across multicast groups and hence provides very high channel quality. With a PDR of 99.999% and 99.52% for high and low LoA, resp., the slow CSI-based RRS scheme

shows a signficantly higher PDR gain as compared to the full CSI-based scheme thanks to the SINR margin.

5.2.2 Sidelink Packet Delay for Large V2N Delay

The QSI allows for a delay-aware RRS as it provides the current queueing delay information on packets in the transmission buffer of each multicast group. In particular, the number of RBs allocated to a multicast group depends on the QSI as well as the channel quality (hence MCS) of each RB. However, in case of a significant delay on V2N links, the QSI perceived at the scheduler placed in the central cloud server (by means of QSI collection from V-UEs) may differ from the actual QSI available at the transmitter V-UE of the multicast groups. This may influence the delay performance of the centralized schemes. To this end, a virtual queue is implemented at the scheduler to track QSI for each multicast group across CAD groups without requiring a need for explicit QSI reporting from V-UEs which anyway would have been outdated in case of large V2N delays. In case of PDT, the QSI can be accurately tracked by knowing the periodicity of the respective data traffic of each multicast group. However, NPDT is expected to pose a challenge since the QSI cannot be accurately tracked at the scheduler.

Fig. 5.8 shows the packet delay performance of the centralized schemes where all the multicast group transmissions are considered to represent PDT. The full CSI-based scheme shows highly desirable 2-5 ms packet delay performance as the ideal case since there is no extra SINR margin allowing more multicast group transmissions to be scheduled at each time slot. However, since it fails to meet the reliability requirements as shown in Fig. 5.7, it is not suitable for CAD when the V2N delay is large. On the other hand, an orthogonal scheme provides ideal reliability, but its packet delay performance is observed to be as high as 395 ms even at the 10th percentile of the delay CDF for high LoA as it assigns too few RBs due to the or-

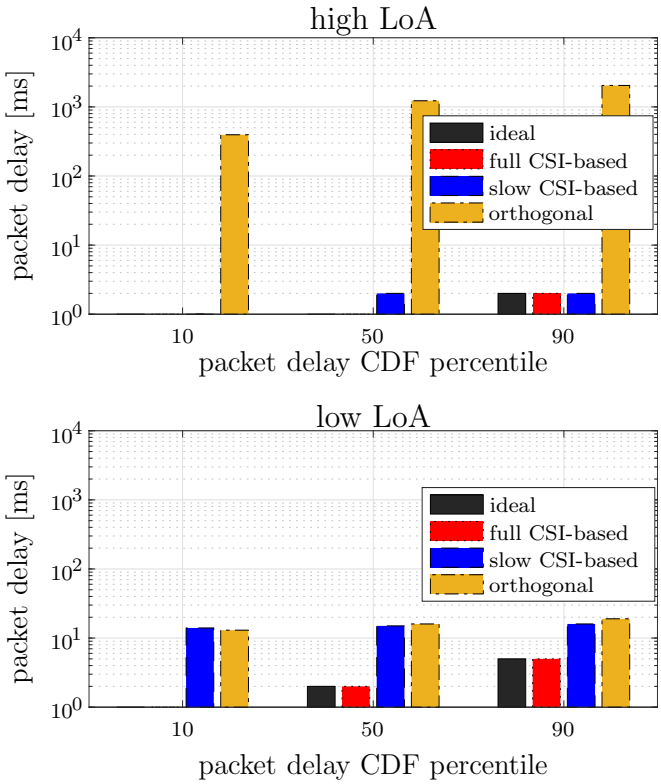


Figure 5.8: Packet delay performance for different centralized schemes for PDT.

thogonal RB allocation. For cases meeting the reliability requirements, it is observed that the slow CSI-based scheme packet delays of 2 ms and 16 ms for high and low LoA, resp., are even at the 90th percentile of the delay CDF, which is well within the 5G requirement [12] of 10 ms and 25 ms for the respective cases.

By contrast, it is challenging for the scheduler to track the NPDT packet generation at the transmitter V-UE of each multicast group accurately. In the evaluation, the scheduler is considered to maintain a virtual queue by following a Bernoulli process with a mean packet arrival rate of $1/T_{\mathrm{pkt},m}$ as described in Sect. 2.3. Fig. 5.9 shows the packet delay performance of the centralized schemes where all the multicast group transmissions are considered to be NPDT. The orthogonal scheme performance for NPDT is observed to be similar to that of PDT and remains unsuitable for CAD due to a significantly higher packet delay. Both full CSI and slow CSI-based schemes show significantly higher packet delay for NPDT. In particular, the

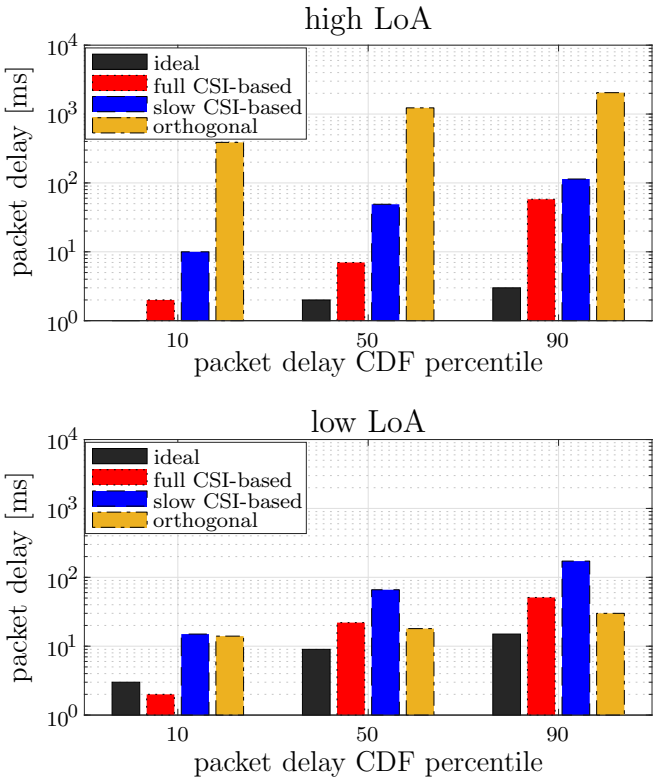


Figure 5.9: Packet delay performance for different centralized schemes for NPDT.

slow CSI-based scheme shows 49 ms and 66 ms packet delay (utility of 0) for high and low LoA, resp., at the median of the packet delay CDF which are significantly above the 5G requirements for the platooning use case. The high packet delay renders both the proposed centralized schemes not suitable for CAD in case of NPDT when the V2N delay is large. This performance degradation is because the multicast groups may not always get sufficient resources from the scheduler in a timely manner since the QSI available at the scheduler does not mirror the actual QSI. Also, there may be a significant resource wastage because RBs may be assigned to multicast groups with empty queues (while virtual queues being non-empty). As a remedy, by bringing the scheduler closer to the V-UEs, the accurate QSI can be made available at the scheduler by collecting the actual QSI with significantly low control plane delay. In this regard, the three-stage resource allocation can be advantageous since the scheduler is at most one hop-away on V2V links from the V-UEs.

5.3 Semi-Centralized Three-Stage Resource Scheduling

5.3.1 Inter-Cluster RB-Reuse Impact on Packet Delay

The three-stage resource allocation offers high tunability by means of two slow CSI measurement thresholds $\bar{\psi}_{\text{ctrl}}$ and $\bar{\psi}_{\text{I}}$ that define bounds on relative control plane reachability among and data plane interference within, resp., multicast groups on V2V links. In the simulation, $\bar{\psi}_{\text{ctrl}}$ is assumed to be $-75\,\text{dBm}$, and upon clustering, the V-UEs in a cluster are assumed to be reachable to each other for intra-cluster RRS. Whereas $\bar{\psi}_{\text{I}}$ enables the scheme to tackle the trade-off between enhancing spectrum utilization by efficient reuse of sidelink RBs among clusters and limiting the inter-cluster interference to an acceptable level. That is, by varying $\bar{\psi}_{\text{I}}$ the RB-reuse and hence the inter-cluster interference can be influenced.

The CDF of the SINR for the three-stage resource allocation scheme for four different values of $\bar{\psi}_{\text{I}}$, namely $-80\,\text{dBm}$, $-95\,\text{dBm}$, $-110\,\text{dBm}$ and $-125\,\text{dBm}$, is shown in Fig. 5.10 for PDT. It can be observed that the tunability offered by the RB-pool allocation allows to achieve different SINR performances. The lower $\bar{\psi}_{\text{I}}$, the lower are the RB reuse and the interference. At $\bar{\psi}_{\text{I}} = -125\,\text{dBm}$, there is almost no RB reuse possible in the considered road segment and hence the SINR performance is similar to the one of the orthogonal scheme shown in Fig. 5.6. Whereas, $\bar{\psi}_{\text{I}} = -110\,\text{dBm}$ allows for RB reuse across multicast group clusters that are located mostly away from the center at the opposide side of the road seqment. For clusters that are located at or near to the center of the road segment, an orthogonal RB-pool is used resulting in different SINR regions as observed in the CDF of the SINR.

Since $\bar{\psi}_{\text{I}}$ influences the RB reuse and the inter-cluster interference, it affects the size of the RB-pool allocated to each cluster and the channel quality for mul-

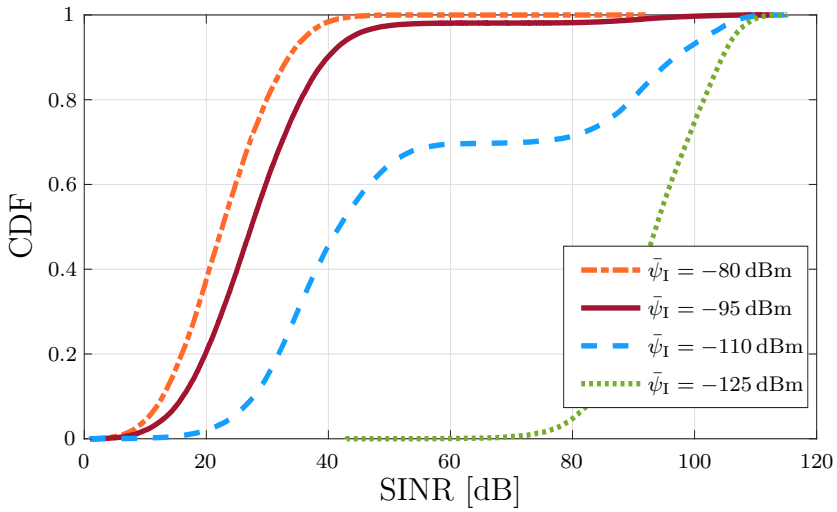


Figure 5.10: SINR performance for different inter-cluster interference bounds $\bar{\psi}_{\mathrm{I}}$ for PDT.

ticast groups and hence the MCS selection. Consequently, $\bar{\psi}_{\mathrm{I}}$ impacts the packet delay performance on V2V links. Fig. 5.11 shows the packet delay performance of the three-stage resource allocation scheme for the aforementioned four different $\bar{\psi}_{\mathrm{I}}$ bounds assuming PDT. It is observed that even though the configuration with $\bar{\psi}_{\mathrm{I}} = -125\,\mathrm{dBm}$ assigns mostly orthogonal RB-pools, it provides packet delays of 5 ms and 18 ms for high and low LoA, resp., at the 90th percentile of the packet delay CDF in contrast to the poor packet delay performance of the centralized orthogonal scheme presented in Fig. 5.8. This is due to the ability of the three-stage resource allocation scheme to acquire and utilize valid full CSI and QSI in a dynamic QoS-aware RRS at the vehicle level. The other values of $\bar{\psi}_{\mathrm{I}}$ lead to highly desirable 1-3 and 1-11 ms packet delays for high and low LoA, respectively. In the following evaluation, $\bar{\psi}_{\mathrm{I}}$ is considered to be $-95\,\mathrm{dBm}$.

To evaluate the ability of the three-stage resource allocation scheme to flexibly adapt to adapt to different bandwidths in order to cater to CAD applications of different delay requirements on V2V links, Fig. 5.12 shows a packet delay performance comparison for different bandwidths where each multicast group is considered to perform PDT transmissions. Unlike Fig. 5.3, this packet delay evaluation does not

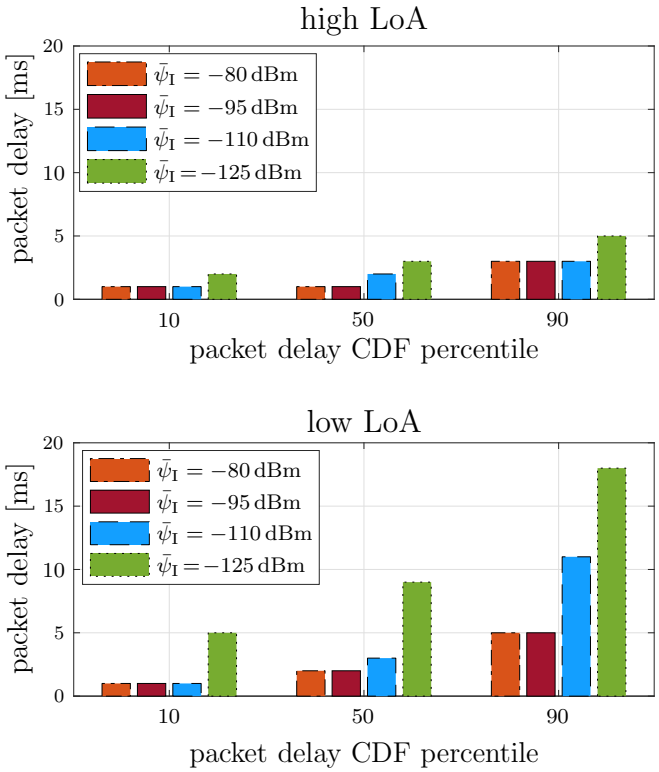


Figure 5.11: Packet delay performance for different inter-cluster interference bounds $\bar{\psi}_{\mathrm{I}}$ in a three-stage resource allocation for PDT.

consider a bandwidth of 1 MHz since 5 RBs may not be sufficient to allocate at least 1 RB to each cluster unless the $\bar{\psi}_{\mathrm{I}}$ is chosen to be significantly higher at the risk of poor V2V links. At the 90th percentile of the packet delay CDF, desirable 4 ms and 7 ms packet delays are observed for high and low LoA, resp., even when the bandwidth is reduced to as low as 2 MHz. Also, the evaluation shows that the scheme is able to flexibly allocate resources considering different levels of packet delays for platooning with high and low LoA.

5.3.2 Sidelink Reliability against Centralized Schemes

Fig. 5.13 shows the CDF of the SINR on V2V links involved in multicast group PDT transmissions for the three-stage resource allocation in comparison with the ideal and the slow CSI-based RRS schemes for both high and low LoA. The SINR

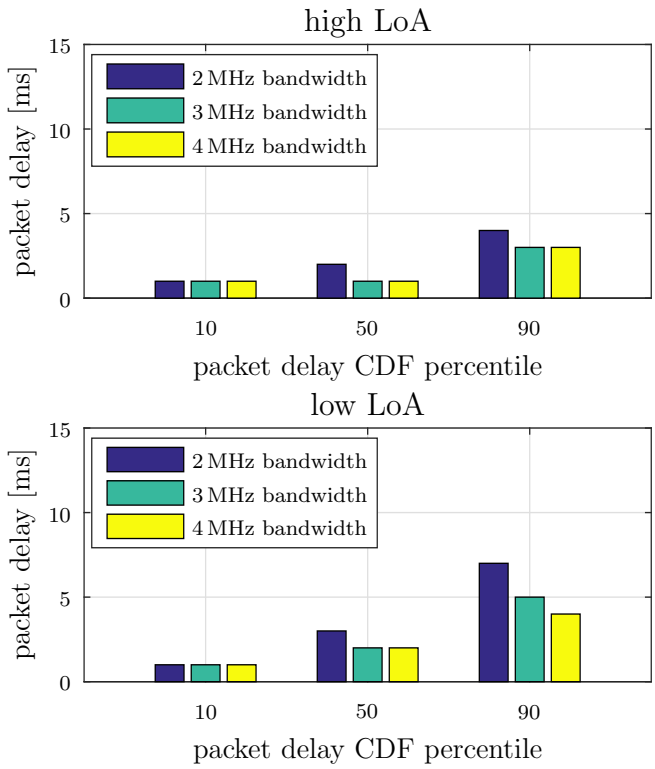


Figure 5.12: Packet delay performance of three-stage RRS for different bandwidths and PDT.

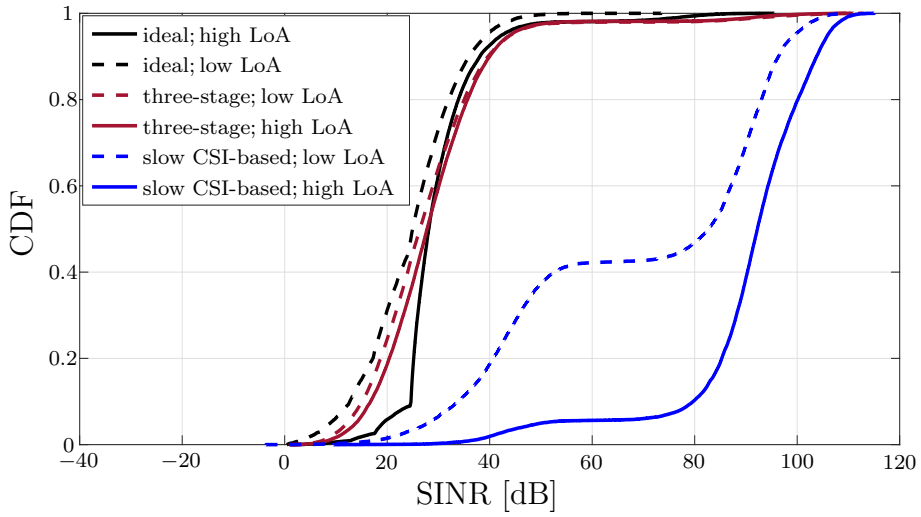


Figure 5.13: SINR performance for different RRS schemes and PDT.

performance of the three-stage resource allocation for high and low LoA is observed to be not significantly different (at most about 1.5 dB) since the inter-cluster RB-pool assignment allows for the same level of inter-cluster interference (because of the same value $\bar{\psi}_{\mathrm{I}}$) irrespectively of the QoS requirements of multicast groups in the cluster. The three-stage scheme shows a slightly higher SINR performance as compared to the ideal case for low LoA, in particular, about 2.5 . . . 4 dB up to the 90th percentile of the CDF since the intra-cluster RRS takes into account the worst-case inter-cluster interference. On the other hand, the ideal case outperforms the three-stage scheme below the median of the CDF as it has the advantage of RRS across all multicast groups. The slow CSI-based centralized RRS significantly outperforms both ideal and three-stage schemes due to the SINR margin used during the RRS. However, a deviation from the ideal case may adversely impact the delay performance as the higher SINR may be the result of a lower RB reuse and hence a lesser number of RBs available for allocation. A lower SINR (with a usage of potentially lower-order MCSs) may result from the fact that the inter-multicast group interference cannot be differentiably limited for high LoA in the three-stage resource allocation as opposed to the case for centralized RRS schemes. In this regard, the three-stage scheme shows a good SINR performance as the deviation from the ideal case is not large.

Fig. 5.14 shows the PDR performance of multicast transmissions associated with both high and low LoA for the three-stage resource allocation in comparison with the centralized RRS schemes for PDT. The three-stage scheme is observed to outperform the slow CSI-based centralized RRS and achieves the ideal PDR of 100%. Since the CH of each cluster collects full CSI from the respective member V-UEs via one hop-away V2V links, the intra-cluster scheduler has the perfect knowledge of the CSI and the QSI. This allows the scheduler to consider the worst-case SINR by computing the worst-case inter-cluster interference to perform link adaptation (i.e.,

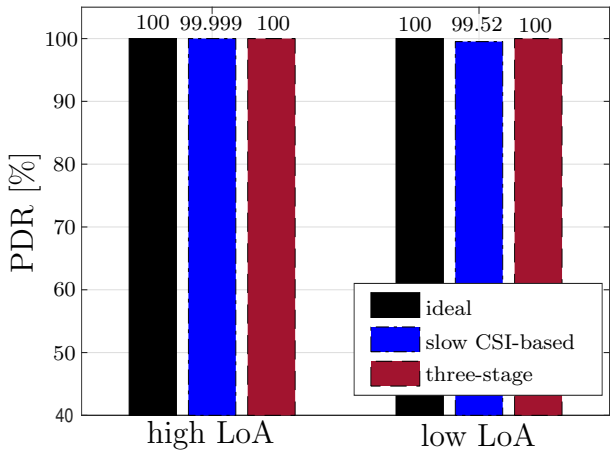


Figure 5.14: PDR performance for different RRS schemes and PDT.

MCS allocation) accordingly to achieve the ideal PDR performance.

5.3.3 Sidelink Packet Delay against Centralized Schemes

The packet delay performances for PDT and for NPDT of the three-stage resource allocation scheme in comparison with the ideal and slow CSI-based RRS schemes are shown in Fig. 5.15 and Fig. 5.16, respectively. Fig. 5.15 shows a highly desirable $1 \ldots 5$ ms packet delay performance for the three-stage resource allocation which is only 1 ms inferior to the ideal case at the 90th percentile of the delay CDF for high LoA. In case of low LoA, the three-stage scheme significantly outperforms the slow CSI-based scheme. In particular, even at the 90th percentile of the packet delay CDF, the three-stage scheme is observed to provide 5 ms (utility of 0.999) in comparison to 16 ms (utility of 0.78) packet delays of the slow CSI-based scheme. The performance gain of the three-stage resource allocation is mainly owed to the resource demand-aware RB-pool allocation, instantaneous QSI collection by the CH over low latency one hop-away V2V links and QoS-aware intra-cluster RRS based on full CSI without considering any SINR margin. Similarly, in case of NPDT, the three-stage scheme shows desirable $1 \ldots 5$ ms and $3 \ldots 8$ ms delays for high and low LoA, resp., as observed in Fig. 5.16.

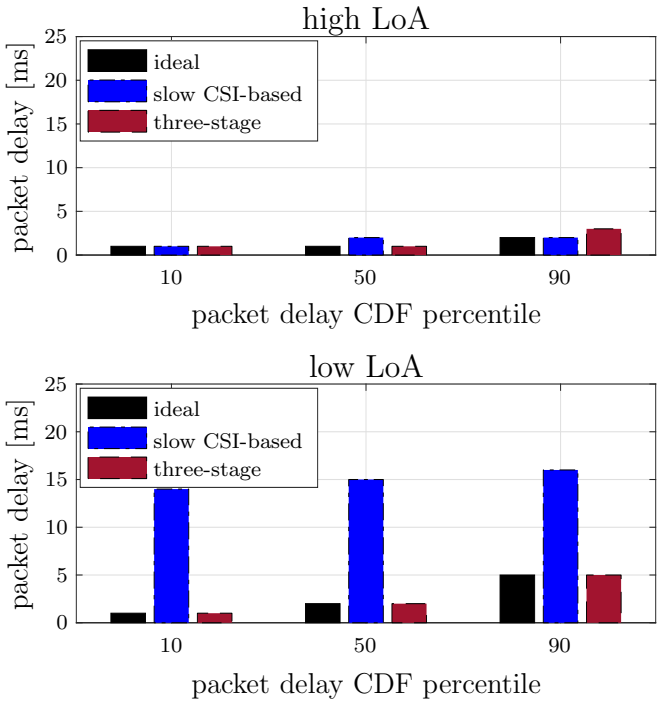


Figure 5.15: Packet delay performance for different RRS schemes for PDT.

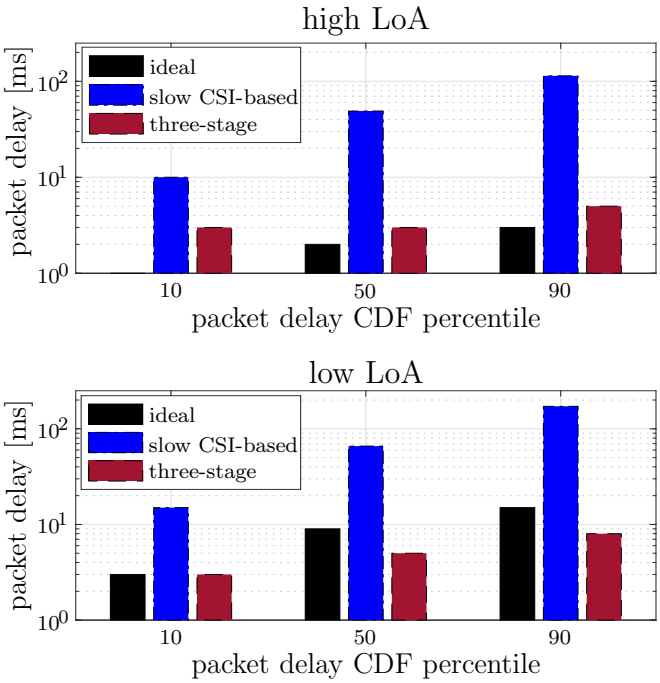


Figure 5.16: Packet delay performance for different RRS schemes for NPDT.

Even at the 90th percentile of packet delay CAD, for high LoA, the scheme shows a desirable 5 ms delay compared to the ideal 3 ms. For low LoA, the three-

stage scheme achieves a delay of 8 ms in comparison to the ideal 15 ms. Note that the ideal case maximizes the utility across all multicast groups and hence it penalizes the delay-tolerant low LoA to gain the performance for high LoA. Since the three-stage scheme aims to maximize the aggregated utility within each cluster and not across all the multicast groups, even though it outperforms the ideal case, it compromises the delay performance for high LoA in comparison to the ideal case. Furthermore, the three-stage scheme significantly outperforms the slow CSI scheme since it can acquire the actual QSI over one hop-away V2V links and does not employ an SINR margin.

It is noted that the slow CSI-based centralized RRS scheme and the three-stage RRS scheme reduce the signaling overhead significantly on V2N links by at least T_{c} times compared to the full CSI-based centralized RRS scheme since the V-UEs need to report CSI to the centralized RRM entity only at an interval of T_{c} as opposed to reporting CSI in every time slot. However, the three-stage resource allocation imposes signaling overhead (which is saved on V2N links) on V2V links since, in turn, all cluster member V-UEs report CSI and QSI to the CH in every time slot.

Chapter 6

Conclusions

6.1 Summary and Results

This thesis discusses sidelink radio resource allocation for multicast group-based V2V communications to support CAD use cases. Utility functions are used to capture QoS requirements of CAD that span a wide range with different levels of elasticity reflecting the performance tolerance of CAD use cases depending on the level of criticality, LoA and cost of resulting suboptimal CAD operation (w.r.t., e.g., fuel or road efficiency). A QoS-dependent utility-based multi-objective RB and MCS allocation optimization problem is formulated for CAD sidelink multicast group communication with the aim to maximize the aggregated utility (hence, the CAD operation efficiency) across multicast groups considering reliability constraints and a half-duplex limitation. Moreover, centralized and semi-centralized RRS solutions are proposed and evaluated using a 5G-based simulator with a realistic vehicular mobility for a platooning use case considering two high and low LoA.

Recognizing the need for sidelink RRM in an operator-independent manner to enable sidelink V2V interactions for CAD and interference coordination among V-UEs in a multi-operator environment and allowing resource control by vertical partners such as automotive organizations, a cloud-enabled sidelink RRM framework is introduced. The latter allows for placing the sidelink radio resource scheduler at a

logically centralized cloud server that can be accessed by V-UEs across MNOs using V2N links and for programmability by third-party users (e.g., vertical partners) or applications via APIs. Exploiting the framework to solve the sidelink resource allocation problem, a full CSI-based centralized QoS-aware RRS scheme is proposed. With full knowledge of CSI and QSI, the scheme flexibly adapts the available bandwidth to cater CAD applications of different packet delay requirements and offers high reliability and low latency when the V2N delays is small/close to zero. However, depending on the deployed backhaul topologies and technologies, the V2N delay between the V-UEs and the cloud server could be significantly larger in practice due to a high transport network delay. This large V2N delay results in a higher scheduling delay and consequently in a larger queuing delay which may not be acceptable for critical CAD use cases such as CCA. Furthermore, as the vehicular radio environment is highly dynamic in case of high vehicular mobility, the large V2N delay may render the CSI being available at the cloud server to be invalid. This results in a poor reliability performance since the perceived V2V channel quality may be signficantly different from the actual channel quality at the multicast group. Hence, the scheme may not be suitable for CAD when the V2N delay is large. In addition, this dynamic scheduling performed at the central cloud server results in a high signaling overhead over the V2N link due to the acquisition of instantaneous CSI and QSI from V-UEs. This may not be desirable, in particular, when the V-UEs utilize V2N links offered by MNOs to exchange RRC information between the V-UEs and the cloud server in a pay-per-use manner.

To this end, a slow CSI-based centralized RRS is proposed which requires V-UEs to report only second-order CSI measurements (consisting of slowly varying large-scale channel parameters) at an interval of tens or hundreds of milliseconds as opposed to full CSI measurement acquisition in each time slot and hence reducing the signaling overhead significantly. The scheme can be implemented in a scheduler at

the central cloud entity and is robust against the V2N delay since slow CSI remains almost unchanged during the V2N delay. It ensures reliable V2V transmissions by providing an extra SINR margin to account for a potential SINR performance deviation due to the influence of fast fading which is not captured by the slow CSI measurements. This, however, comes at the cost of resource utilization (i.e., lower RB reuse) where the SINR margin increases drastically for an increasing reliability requirement (e.g., to 39.99 dB for 99.99% reliability) which limits the packet delay performance. Furthermore, similarly to the case of a full CSI-based RRS scheme, a large V2N delay results in a higher scheduling delay and therefore a larger queuing delay. While the RRS can be peformed for PDT and a known periodicity, this is particularly challenging for NPDT that cannot be predicted with high accuracy. Therefore, in case of a large V2N delay, the solution is more suitable for CAD use cases involving PDT.

The impact of the V2N delay in dynamic RRS can be eliminated by placing the scheduler closer to the V-UEs. In this regard, an edge-cloud enabled semi-centralized sidelink RRM framework is presented by extending the cloud-enbled sidelink RRM framework to leverage the VEC capabilities of autonomous vehicles to bring the dynamic scheduler closer to the V-UEs. By exploiting the framework, a semi-centralized three-stage RRS scheme is proposed where the resource allocation problem is split into three modular and tunable functional tasks to reduce scheduling delay and V2N signaling overhead. At the central cloud server, at first, V-UEs are clustered ensuring control plane reachability to enable intra-cluster V2V transmission coordination. Subsequently, each cluster is provided with an RB-pool considering the resource demands and a trade-off between resource utilization and V2V link quality. Finally, intra-cluster resource allocation is performed dynamically at the VEC nodes by collecting CSI and QSI from member V-UEs over a one-hop V2V link to meet the QoS requirements of CAD. The scheme offers high reliability

since it considers the worst-case SINR by computing the worst-case inter-cluster interference using accurate CSI and performs link adaptation (i.e., MCS allocation) accordingly. Also, the scheme supports low-latency CAD V2V communication and offers a high packet-delay performance gain as compared to a slow CSI-based centralized scheme even for NPDT thanks to the resource demand-aware RB-pool allocation, instantaneous QSI collection by the CH over low-latency one-hop away V2V links and QoS-aware intra-cluster RRS based on full CSI without considering any SINR margin. Furthermore and similarly to the case of slow CSI-based centralized RRS, the signaling overhead over V2N is significantly reduced because the V-UEs need to report only the slow CSI at an interval of tens or hundreds of milliseconds. However, since the scheduler collects CSI over V2V links, the scheme imposes a signling overhead (which is saved on V2N links) on V2V links. It is noted that the edge-cloud enabled semi-centralized sidelink RRM framework may be more complex to be deployed than the cloud-based centralized sidelink RRM framework since it involves setting up and maintaining VEC nodes. Yet, it allows the RRM to be split into long-term and dynamic tasks to be performed in a centralized cloud server and VEC nodes, resp., at different time scales to selectively achieve different RRM objectives (e.g., low signaling overhead over V2N links, high reliability and/or low latency on V2V links, etc.) and to effectively serve different CAD use cases.

6.2 Outlook

The following aspects should be considered in potential extensions that can further enhance the sidelink RRM for CAD use cases.

- Enhancements to sidelink CSI acquisition at the V-UEs including the resource allocation for sidelink reference signal transmission: This work assumes that all V-UEs are able to exchange reference signals and the CSI to be available

at the V-UEs for all the associated V2V links. However, the transmission of sidelink reference signals increases the signaling overhead on V2V links such that the resource allocation for sidelink reference signal transmission can be challenging in view of high vehicular mobility and half-duplex constraints. Hence, strategies for sidelink CSI acquisition that capture the accurate channel state with reduced the signaling overhead on V2V links are required.

- Data traffic prediction models for CAD: Accurate prediction of CAD data traffic can reduce the impact of the V2N delay in centralized RRS by allocating the resources beforehand to reduce the scheduling delay. Also, this allows for a proportional RB-pool allocation in case of the three-stage RRS scheme. In this regard, machine learning (ML)-based data driven strategies for CAD data traffic prediction, based on, e.g., the nature of the CAD operation and the events that potentially trigger the operation, can significantly improve the sidelink packet performance, in particular for NPDT.

- Employing reinforcement learning (RL)-based optimal resource allocation algorithms in sidelink RRS: This work proposes simpler heuristic solutions for RRS that are easy to implement. While finding the optimal resource allocation solution may be highly complex and computationally expensive, ML-based solutions may show a potential to solve the dynamic resource allocation problem optimally. In this regard, RL-based agents [69], [70] can potentially be trained to solve the combinatorial sidelink resource allocation optimization problem. Here, an aggregated multicast group utility or resource utilization may be considered as the reward for the action of resource allocation by the agent.

- Predictive QoS to avoid frequent adaptation of CAD operation that may adversely affect the CAD operation objective: CAD applications adapt the CAD operation, e.g., control of the IVD in platooning, based on the QoS perfor-

mance on the V2V link, e.g., to maintain the safety gap. However, frequent IVD changes necessitate dynamic acceleration and deceleration of vehicles that may increase fuel consumption at platoons and hence may defeat the purpose of platooning. Therefore, a CAD application can significantly benefit from the early notification of QoS performance degradation or improvement so that it can plan its vehicle control operation to maximize the CAD operation objective (e.g., fuel saving). Hence, strategies to predict the QoS performance is desirable.

- Transmission power adaptation: This work assumes fixed transmission power for both control and payload data transmissions. Yet, since the transmission power adapation has the potential to significantly improve V2V link quality by keeping the interference to an acceptable level, strategies for optimal transmission power allocation or adaption is beneficial in supporting CAD V2V communication.

Appendix A

SAE Levels of Automation

Table A.1 provides the narrative definition of LoA as per SAE International.

Table A.1: Definition of LoA [2].

SAE level	notion	narrative definition
0	No Automation	full-time performance by the human driver in all aspects of the DDT, even when enhanced by warning or intervention systems
1	Driver Assistance	driving mode-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the DDT
2	Partial Automation	driving mode-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the human driver performs all remaining aspects of the DDT

3	Conditional Automation	driving mode-specific performance by an ADS of all aspects of the DDT with the expectation that the human driver will respond appropriately to a request to intervene
4	High Automation	driving mode-specific performance by an ADS of all aspects of the DDT, even if a human driver does not respond appropriately to a request to intervene
5	Full Automation	full-time performance by an ADS of all aspects of the DDT under all roadway and environmental conditions that can be managed by a human driver

Appendix B

5G Application Layer Support for V2X Services

In Release-16, 3GPP has introduced application layer functional support to V2X services. The vertical application layer (VAL) for V2X services is grouped into the V2X application-specific layer and the VAE layer [55]. The VAE layer offers the VAE capabilities to the V2X application-specific layer utilizing the SEAL services offered by the 3GPP network as shown in Fig. B.1. Here, the V2X application server consists of an V2X application-specific server, a VAE server and SEAL servers. The V2X application-specific server provides the server-side functionalities corresponding to the V2X applications (e.g., platooning server) and utilizes the VAE server for the V2X application layer support functions. Furthermore, the V-UE clients in the autonomous vehicles contain the V2X application-specific client, the VAE client and the SEAL clients.

The V2X applications are linked with vehicular dynamic models and control. They carry out CAD operations, e.g., CLC and platooning, by defining and solving control problems as per the use cases at hand. Furthermore, they are responsible for choosing appropriate LoA. These V2V applications are configured and/or handled by a V2X application-specific server (e.g., platooning server), which may be deployed per geographical area or per service with application-level signaling to the V2X application-specific clients at the V-UEs.

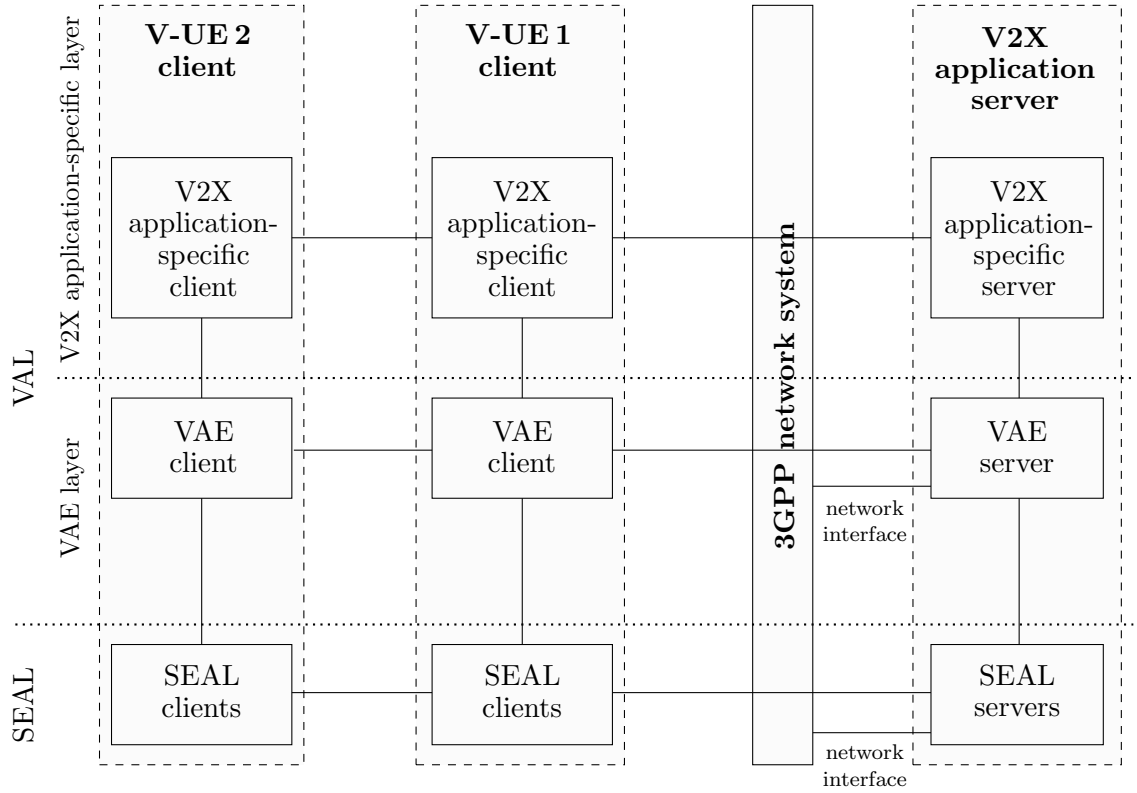


Figure B.1: 5G V2X application layer functional model [55] (reference points between entities not shown explicitly).

A V2X application enabler is offered as API to a V2X application and utilizes SEAL services to allow interaction between the 3GPP network and the V2X application. VAE capabilities provide support for application level location tracking and group management (e.g., creation, join, leave). Furthermore, the VAE server supports monitoring of a 3GPP systems' network situation (including RAN-related resource/traffic situations) for V2X services and QoS for a single vehicle or groups of vehicles (supporting a V2X service and being in proximity) having ongoing sessions. In addition, a V2X application provides service requirements to the 3GPP network system through a VAE.

The SEAL in 3GPP networks offers its services to the vertical application layer to support vertical industry applications (e.g. V2X applications) [56]. The SEAL server(s) may communicate with the underlying 3GPP network systems using the

respective 3GPP interfaces and offer support through APIs to expose SEAL services to the vertical applications. SEAL services include location management, GM (e.g., group creation and deletion), configuration management and NRM (e.g., unicast, multicast resources).

Utilizing the aforementioned application layer support, the interaction between the sidelink RRM and CAD V2X applications can be enabled. A V2X application-specific server can distribute the network QoS requirements of CAD and LoA on rate, reliability and/or delay via a VAE. Based on the CAD coordination control requirement, the application-specific server can also provide information to the network on the V2V transmission mode (e.g., unicast, multicast) and on the group of vehicles involved in the CAD. With this information, the network can become aware of the required vehicle coordination control and can perform sidelink RRM accordingly to maximize the CAD performance. In case of centralized sidelink RRM, the scheduler can be placed at a central server which is co-located with a SEAL server to get the SEAL services.

The semi-centralized sidelink RRM that is presented in Chapter 4 can also be realized exploiting the application layer support as shown in Fig. B.2 for a 3GPP-compliant network. The CM and RC functions, which update vehicle clusters and RB-pool configuration periodically, can be placed at a central server which is co-located with a SEAL server to allow closer interaction with the GM and NRM server functionalities of the SEAL. On the other hand, the resource scheduler can be placed at the VEC being formed by the vehicles in the cluster and interacting with clients for GM and NRM. The output (e.g., clusters and RB-pools) of CM and RC functionalities are conveyed to the RRS through the SEAL's GM and NRM, respectively.

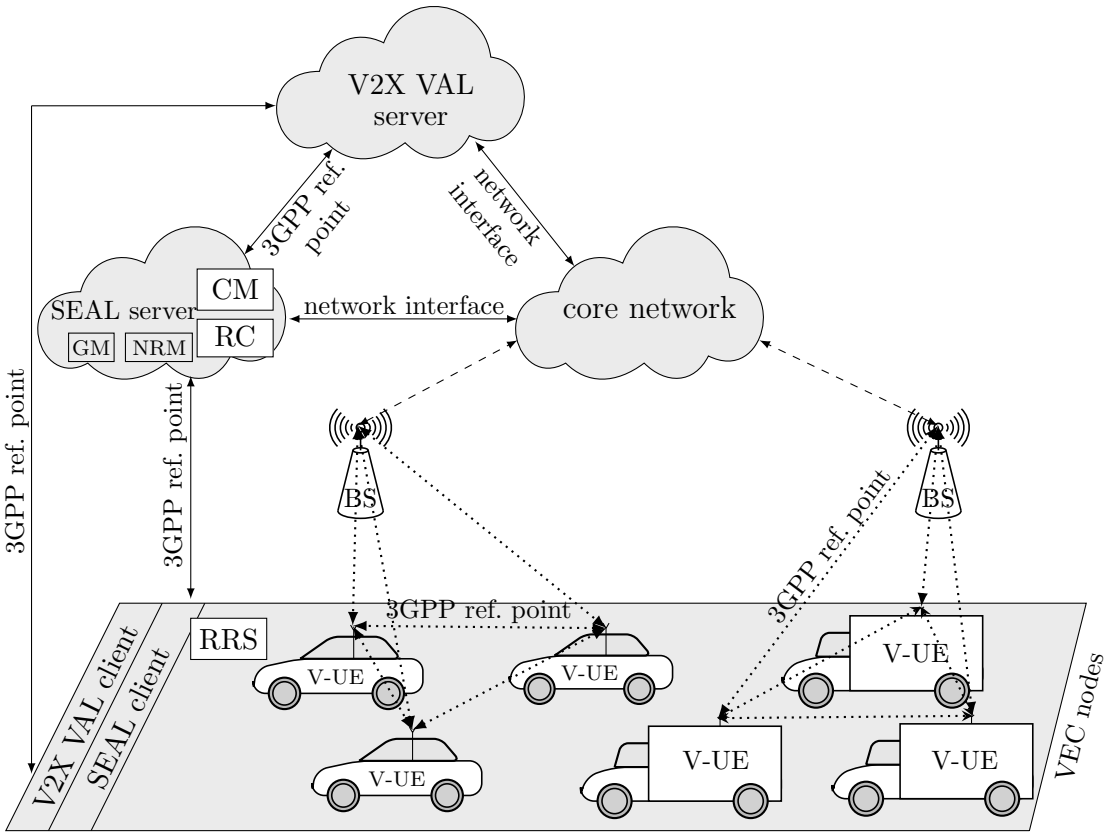


Figure B.2: Edge cloud-enabled sidelink RRM split using a 5G V2X application layer functional architecture.

Appendix C

Glossary

CAD ⇔ A means for vehicles to perform co-operative trajectory planning and control as well as co-operative perception and road environment sensing by leveraging their communication capabilities, p. 2

CAD group ⇔ A set of V-UEs which are participating (or co-operating) to carry out a CAD operation, e.g., platooning. Each CAD group may consist of multiple multicast groups, p. 23

CCA ⇔ A CAD use case in which vehicles coordinate their movements to avoid a collision when one or more involved vehicles are at risk of a collision, and return to their original course when the risk is averted, p. 4

CLC ⇔ A CAD use case in which vehicles coordinate to perform an autonomous lane change operation for one or more of the involved vehicles, p. 3

cloud computing ⇔ On-demand availability and access of computing and storage resources via Internet, p. 11

cloud server ⇔ A server with a pool of large storage and computational resources being available on demand for computation and storage and remotely accessible via Internet, p. 12

control plane ⇔ A blanket term for all signaling (e.g., RRC messages) that controls or manages communication links for payload data transmission, p. 9

edge computing ⇔ A distributed computing paradigm where the computation and storage resources are made available at the edge of the network near to or co-located with the end devices, p. 13

full CSI ⇔ CSI consisting of both small-scale parameters, i.e., fast fading components, and large-scale channel parameters like path loss and shadowing, p. 14

HOL packet delay ⇔ Queing delay associated with the foremost packet (the packet which is waiting for the longest to be dequeued) in the transmission packet queue, p. 31

IVD ⇔ Euclidean distance between the center point of the rear bumper of the leading vehicle and the front bumper of the trailing vehicle, p. 3

MNO ⇔ A wireless communication service provider who owns or controls the necessary elements such as a wireless network infrastructure (e.g., BSs) and radio spectrum being needed to enable wireless communication, p. 9

multicast group ⇔ A unique identifier for a group of V-UEs consisting of a transmitting V-UE and a set of receiving V-UEs for a multicast transmission within a CAD group, p. 18

platooning ⇔ A CAD use case where multiple vehicles travel as a group with reduced IVD where a lead vehicle is followed by the trailing vehicles within the group, p. 3

QoS ⇔ A measure of communication performance experienced by V-UEs, e.g., data rate, packet delay or link availability, p. 5

RAN ⇔ A part of the communication network which connects the devices such as V-UEs to the other part of the network, e.g., to the core network via radio

Appendix D

List of Abbreviations

3GPP 3rd generation partnership project, p. 5

5G fifth generation mobile networks, p. 5

5GAA 5G automotive association, p. 3

ADS automated driving system, p. 1

API application programming interface, p. 12

BLER block-error rate, p. 28

BM benchmark, p. 99

BS base station, p. 6

CAD co-operative automated driving, p. 2

CCA co-operative collision avoidance, p. 4

CDF cumulative distribution function, p. 96

CH cluster head, p. 70

CLC co-operative lane change, p. 3

CM cluster management, p. 69

CPP clique partitioning problem, p. 18

CSI channel state information, p. 14

DDT dynamic driving task, p. 1

eNB eNodeB, p. 8

FIFO first-in first-out, p. 26

FMeC follow-me edge-cloud, p. 16

GM group management, p. 40

gNB gNodeB, p. 8

HOL head-of-line, p. 31

HVC hybrid vehicular edge cloud, p. 17

IEEE Institute of Electrical and Electronics Engineers, p. 9

IP Internet protocol, p. 38

ITS intelligent transportation system, p. 12

IVD inter-vehicular distance, p. 3

KPI key performance indicator, p. 34

LDPC low-density parity-check, p. 96

LoA levels of automation, p. 1

LTE long-term-evolution, p. 8

MAC medium access control, p. 37

MCS modulation and coding scheme, p. 7

MEC multi-access edge computing, p. 16

ML machine learning, p. 121

MNO mobile network operator, p. 9

NPDT non-periodic data traffic, p. 26

NR new radio, p. 8

NRM network resource management, p. 40

NUM network utility maximization, p. 43

OEDR object and event detection and response, p. 2

OFDM orthogonal frequency-division multiplexing, p. 22

OFDMA orthogonal frequency-division multiple access, p. 27

PDR packet delivery ratio, p. 96

PDT periodic data traffic, p. 26

QoS quality-of-service, p. 5

QSI queue state information, p. 36

RAN radio access network, p. 11

RB resource block, p. 8

RC resource configuration, p. 69

RE resource element, p. 28

RL reinforcement learning, p. 121

RRC radio resource control, p. 9

RRM radio resource management, p. 7

RRS radio resource scheduling, p. 7

RSU roadside unit, p. 16

SAE society of automotive engineers, p. 1

SCS subcarrier spacing, p. 95

SEAL service enabler architecture layer, p. 40

SINR signal-to-interference-plus-noise ratio, p. 28

SPS semi-persistent scheduling, p. 99

TB transport block, p. 96

V-UE vehicle user-equipment, p. 6

V2I vehicle-to-infrastructure, p. 6

V2N vehicle-to-network, p. 12

V2V vehicle-to-vehicle, p. 4

V2X vehicle-to-everything, p. 8

VAE V2X application enabler, p. 40

VAL vertical application layer, p. 125

VEC vehicular edge computing, p. 16

VMC vehicle motion control, p. 2

Appendix E

List of Symbols

Latin and Caligraphic Symbols

a	index of an MCS with $a \in \Omega_{1,A}$, p. 28
A	number of MCSs, p. 28
B_m	packet size in bits of each packet that is generated by m^{th} multicast group, p. 26
c	index of a V-UE cluster with $c \in \Omega_{1,C}$, p. 80
c'	index of a V-UE cluster with $c' \in \Omega_{1,C}$, $c' \neq c$, p. 80
C	number of V-UE clusters, p. 79
d	index of a V-UE in the set of V-UEs with $d \in \Omega_{1,D}$, p. 23
$d_{\text{tx},m}$	transmitter V-UE index for multicast group of index m with $d_{\text{tx},m} \in \mathcal{D}_m$, $d_{\text{tx},m} \notin \mathcal{D}_{\text{rx},m}$, p. 24
$\tilde{d}$	index of a V-UE in the set of V-UEs with $\tilde{d} \in \Omega_{1,D}$, $\tilde{d} \neq d$, p. 23
D	number of V-UEs, p. 22
$\mathcal{D}_m$	index set of member V-UEs of m^{th} multicast group with $\mathcal{D}_m \subseteq \Omega_{1,D}$, p. 24
$\mathcal{D}_{\text{rx},m}$	index set of recipient V-UE for m^{th} multicast group with $\mathcal{D}_{\text{rx},m} \subset \mathcal{D}_m$, p. 24
$\mathrm{E}_{\mathrm{V}_i,\mathrm{V}_j}$	edge between vertices V_i and V_j on graph G with $\mathrm{E}_{\mathrm{V}_i,\mathrm{V}_j} \in \mathcal{E}$, p. 84
$\mathrm{E}^{\backprime}_{\mathrm{V}^{\backprime}_{i^{\backprime}},\mathrm{V}^{\backprime}_{j^{\backprime}}}$	edge between vertices $\mathrm{V}^{\backprime}_{i^{\backprime}}$ and $\mathrm{V}^{\backprime}_{j^{\backprime}}$ on graph $\mathrm{G}^{\backprime}$ with $\mathrm{E}^{\backprime}_{\mathrm{V}^{\backprime}_{i^{\backprime}},\mathrm{V}^{\backprime}_{j^{\backprime}}} \in \mathcal{E}^{\backprime}$, p. 86

$\mathcal{E}$	edge set of graph G, p. 84
$\mathcal{E}^{\backprime}$	edge set of graph G$^{\backprime}$, p. 86
f_c	RB-pool allocation outcome for c^{th} cluster, p. 81
$\mathbf{f}$	RB-pool allocation outcome vector, p. 81
$\mathcal{F}_{c,n}$	index set of multicast groups with non-empty queues at time slot n in c^{th} cluster, p. 82
$\mathcal{F}_n$	index set of multicast groups with non-empty queues at time slot n, p. 45
$\check{\mathcal{F}}_{c,n}$	index set of multicast groups with non-empty queues at time slot n in c^{th} cluster that are nominated for resource allocation, p. 89
$\check{\mathcal{F}}_n$	index set of multicast groups with non-empty queues at time slot n that are nominated for resource allocation, p. 48
G	inter-CAD group control plane reachability graph, p. 84
G$^{\backprime}$	resource-conflict graph, p. 86
$h_{\mathrm{L}_{d,\tilde{d}},k,n}$	normalized small-scale fading random variable over V2V link $\mathrm{L}_{d,\tilde{d}}$ in k^{th} RB at time n, p. 29
$H_{\mathrm{L}_{d,\tilde{d}},k,n}$	channel gain over the V2V link $\mathrm{L}_{d,\tilde{d}}$ in k^{th} RB at time n, p. 29
$\bar{H}_{\mathrm{L}_{d,\tilde{d}}}$	slow CSI over the V2V link $\mathrm{L}_{d,\tilde{d}}$ available at an arbitrary time slot, p. 77
$\bar{H}_{\mathrm{L}_{d,\tilde{d}},n}$	latest acquired slow CSI for V2V link $\mathrm{L}_{d,\tilde{d}}$ available at the scheduler at time slot n, p. 55
i	vertex index of graph G, p. 84
$i^{\backprime}$	vertex index of graph G$^{\backprime}$, p. 86
$I\left(\mathrm{E}^{\backprime}_{\mathrm{V}^{\backprime}_{i^{\backprime}},\mathrm{V}^{\backprime}_{j^{\backprime}}}\right)$	inter-multicast group worst-case interference between the $i^{\backprime\text{th}}$ and $j^{\backprime\text{th}}$ clusters, p. 86
j	vertex index of graph G with $j \neq i$, p. 84
$j^{\backprime}$	vertex index of graph G$^{\backprime}$ with $j^{\backprime} \neq i^{\backprime}$, p. 86
k	index of RBs with $k \in \Omega_{1,K}$, p. 28

K	number of RBs, p. 27
K_{RE}	number of symbols (or REs) in each RB, p. 27
$\mathcal{K}_{\mathrm{V}^{\setminus}_{i^{\setminus}}}$	RB-pool attribute of vertex $\mathrm{V}^{\setminus}_{i^{\setminus}}$, p. 86
$\mathrm{L}_{d,\tilde{d}}$	V2V link between d^{th} transmitter and $\tilde{d}^{\mathrm{th}}$ receiver V-UEs with $\left(d,\tilde{d}\right) \in \Omega^2_{1,D}$, $\tilde{d} \neq d$, p. 23
$\mathcal{L}$	set of all V2V links, p. 23
m	index of a multicast group with $m \in \Omega_{1,M}$, p. 24
m'	index of a multicast group with $m' \in \Omega_{1,M}$, $m' \neq m$, p. 78
M	number of multicast groups across CAD groups, p. 24
$\mathbf{M}$	multicast group cluster matrix, p. 78
$\hat{\mathbf{M}}_{\mathrm{opt}}$	optimal multicast group cluster matrix, p. 79
$\mathcal{M}_{\mathrm{CAD},x}$	index set of multicast groups within x^{th} CAD group with $\mathcal{M}_{\mathrm{CAD},x} \subseteq \Omega_{1,M}$, p. 24
$\mathcal{M}_{\mathrm{clust},c}$	index set of multicast groups in c^{th} cluster with $\mathcal{M}_{\mathrm{clust},c} \subseteq \Omega_{1,M}$, p. 80
n	index of time slot, p. 27
N	number of time slots, p. 27
P_{ctrl}	transmission power for control signaling in an RB at a time slot, p. 30
P_{data}	transmission power for data signaling in an RB at a time slot, p. 30
$P_{\mathrm{L}_{d,\tilde{d}},k,n}$	transmission power over link $\mathrm{L}_{d,\tilde{d}}$ in k^{th} RB at time n, p. 29
$q_{m,\delta,n}$	number of untransmitted bits of the packet enqueued at the beginning of time slot $n-\delta$, $\delta \in \mathbb{N}_0$, for m^{th} multicast group, p. 27
$Q_{m,n}$	number of bits in the queue at the beginning of time slot n for m^{th} multicast group, p. 27
r	transmission rate for an arbitrary multicast group transmission, p. 32
$r_{m,n}(\mathbf{\Pi})$	transmission rate (or rate for short) at time slot n for m^{th} multicast group for resource allocation $\mathbf{\Pi}$, p. 31
$R(a)$	code rate of a^{th} MCS, p. 28

$S(a)$ number of coded bits per symbol in a^{th} MCS, p. 28

T_{c} slow CSI measurement reporting time interval, p. 54

$T_{\text{pkt},m}$ packet generation time interval in slots at the transmitter of m^{th} multicast group in case of PDT, p. 26

T_{sl} duration of time slot, p. 27

T_{V2N} fixed communication delay between a V-UE and the remote cloud server, p. 23

$u_{\text{del},m}(\tau)$ single criterion utility function depending on delay τ for m^{th} multicast group, p. 32

$u_{\text{rate},m}(r)$ single criterion utility function depending on rate r for m^{th} multicast group, p. 32

$u_{\text{rel},m}(\rho)$ single criterion utility function depending on reliability ρ for m^{th} multicast group, p. 32

$\mathbf{u}$ row vector of ones, p. 82

$U_m(r,\rho,\tau)$ utility function depending on rate r, reliability ρ and delay τ for m^{th} multicast group, p. 32

V_i i^{th} vertex of graph G with $\text{V}_i \in \mathcal{V}$, p. 84

$\text{V}'_{i'}$ i'^{th} vertex of graph G$'$ with $\text{V}'_{i'} \in \mathcal{V}'$, p. 86

$\mathcal{V}$ vertex set of graph G, p. 84

$\mathcal{V}'$ vertex set of graph G$'$, p. 86

$W_{\text{clust},c,c'}$ averaged inter-cluster affinity weight for c^{th} and c'^{th} cluster, p. 85

$W_{\text{E}}\left(\text{E}_{\text{V}_i,\text{V}_j}\right)$ affinity among multicast groups across two CAD groups corresponding to vertices connected by $\text{E}_{\text{V}_i,\text{V}_j}$, p. 84

$W_{\text{V}}(\text{V}'_{i'})$ demand attribute of vertex $\text{V}'_{i'}$, p. 86

$\mathbf{W}$ inter-multicast group weight matrix, p. 78

x index of a CAD group with $x \in \Omega_{1,X}$, p. 24

x' index of a CAD group with $x' \in \Omega_{1,X}$, $x' \neq x$, p. 24

X number of CAD groups, p. 24

y_c number of RBs allocated to c^{th} cluster, p. 81

$\mathbf{y}$ RB-pool size vector, p. 81

$\mathbf{Y}$ RB-pool allocation matrix, p. 81

$\hat{\mathbf{Y}}_{\text{opt}}$ optimal RB-pool allocation matrix, p. 82

$\mathbf{Z}$ V2V control plane reachability matrix, p. 78

Greek Symbols

α_m confidence level of fast fading-induced outage probability for m^{th} multicast group, p. 56

$\beta(a, \gamma)$ BLER for a^{th} MCS given an SINR value of γ over a V2V link, p. 28

γ SINR of an arbitrary V2V link, p. 28

$\gamma_{\mathrm{L}_{d,\tilde{d}},k,n}(\mathbf{\Pi})$ SINR over V2V link $\mathrm{L}_{d,\tilde{d}}$ at time slot n in k^{th} RB for resource allocation $\mathbf{\Pi}$, p. 30

$\gamma_{m,k,n}(\mathbf{\Pi})$ channel quality for m^{th} multicast group transmission at time slot n in the k^{th} RB for resource allocation $\mathbf{\Pi}$, p. 30

$\gamma_{\text{thr}}(a)$ channel quality requirement for an arbitrary multicast group given the a^{th} MCS, p. 96

$\gamma_{\text{thr},m}(a)$ channel quality requirement for m^{th} multicast group given the a^{th} MCS, p. 46

$\bar{\gamma}_{\mathrm{L}_{d,\tilde{d}},k,n}(\mathbf{\Pi})$ SINR computed based on only slow CSI for V2V link $\mathrm{L}_{d,\tilde{d}}$ in k^{th} RB at time slot n for resource allocation $\mathbf{\Pi}$, p. 55

$\bar{\gamma}_{m,k,n}(\mathbf{\Pi})$ slow CSI-based channel quality for m^{th} multicast group transmission at time slot n in k^{th} RB for resource allocation $\mathbf{\Pi}$, p. 55

$\dot{\gamma}_{m,k,n}(\mathbf{\Pi})$ worst-case channel quality at m^{th} multicast group in k^{th} RB at time slot n in three-stage resource allocation $\mathbf{\Pi}$, p. 82

Γ_c aggregated resource demand of the multicast groups within c^{th} cluster, p. 81

Δ_m length of packet queue in packets at the transmitter of m^{th} multicast group, p. 26

ϵ steepness parameter of sigmoid utility function, p. 95

ζ center parameter of sigmoid utility function, p. 95

$\eta_{\mathrm{L}_{d,\tilde{d}},n}$ distance-dependent path loss over V2V link $\mathrm{L}_{d,\tilde{d}}$ at time n, p. 29

$\iota_{\tilde{d},k,n}$ worst-case interference power at $\tilde{d}^{\mathrm{th}}$ V-UE from all other V-UE clusters at time slot n in the k^{th} RB, p. 82

$\bar{\iota}_{\mathrm{clust},c,c'}$ slow CSI-based worst-case inter-cluster interference metric between the c^{th} and c'^{th} clusters, p. 80

$\bar{\iota}_{m,m'}$ slow CSI-based interference metric to account for the interference at the m^{th} multicast group caused by the m'^{th} multicast group, p. 80

μ_d index of multicast group in which d^{th} V-UE is the transmitter with $\mu_d \in \Omega_{1,M}$, p. 25

ξ_m SINR margin for m^{th} multicast group transmission, p. 56

$\xi_{\mathrm{thr},m}(\alpha)$ minimum SINR margin needed to achieve reliability requirement of m^{th} multicast group given confidence level α, p. 56

$\mathbf{\Pi}$ resource allocation matrix, p. 28

$\hat{\mathbf{\Pi}}_{\mathrm{opt},n}$ optimal resource allocation at time slot n, p. 45

$\hat{\mathbf{\Pi}}_{\mathrm{opt},n,c}$ intra-cluster resource allocation within c^{th} cluster, p. 83

$\grave{\mathbf{\Pi}}_{\mathrm{opt},n}$ optimal slow CSI-based resource allocation at time slot n, p. 57

ρ reliability for an arbitrary multicast group transmission, p. 32

$\rho_{m,k,n}$ reliability for V2V transmission within m^{th} multicast group in k^{th} RB at time slot n, p. 31

$\rho_{m,k,n}(\mathbf{\Pi})$ reliability for V2V transmission within m^{th} multicast group in k^{th} RB at time slot n for resource allocation $\mathbf{\Pi}$, p. 31

$\rho_{m,n}(\mathbf{\Pi})$ minimum reliability achieved by V2V transmissions within m^{th} multicast group at time slot n for resource allocation $\mathbf{\Pi}$, p. 31

$\rho_{\mathrm{req},m}$ reliability requirement for V2V transmission for m^{th} multicast group, p. 46

σ^2 power spectral density of a standard zero-mean Gaussian thermal noise process, p. 30

τ delay for an arbitrary multicast group transmission, p. 32

$\tau_{m,n}(\mathbf{\Pi})$ HOL packet delay at the end of time slot n (referred to as delay for short) for m^{th} multicast group for resource allocation $\mathbf{\Pi}$, p. 31

$\hat{\Upsilon}_m$ estimated resource demand for m^{th} multicast group, p. 77

$\bar{\psi}_{\text{ctrl}}$ channel quality threshold for the received power on the control V2V link below which multicast groups cannot coordinate their transmissions with each other, p. 77

$\bar{\psi}_{\text{I}}$ a threshold for allowed slow CSI-based inter-cluster interference, p. 77

$\Omega_{a,b}$ index set $\{n \in \mathbb{N}_0 : a \leq n \leq a + b - 1\}$ of cardinality b defined for $a \in \mathbb{N}_0$, $b \in \mathbb{N}$, p. 23

Appendix F

List of Publications

- P. Keshavamurthy, E. Pateromichelakis, D. Dahlhaus, and C. Zhou, "Edge Cloud-Enabled Radio Resource Management for Co-Operative Automated Driving", *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 7, pp. 1515–1530, 2020.

- P. Keshavamurthy, E. Pateromichelakis, D. Dahlhaus, and C. Zhou, "Resource Scheduling for V2V Communications in Co-Operative Automated Driving", in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

- P. Spapis, P. Keshavamurthy, and C. Zhou, "User Equipment for V2V Communications", pat. PCT/EP2018/077301, Apr. 16, 2020.

- S. Ayaz, D. Medina, P. Keshavamurthy, and C. Zhou, "Devices and Methods for Cloud-based Sidelink Scheduling and Base Station Interface Therefor", U.S. Patent 16/700,683, Dec. 2, 2019.

- S. Ayaz, D. Medina, P. Keshavamurthy, and M. Dillinger, "Devices and Methods for Communication in a Wireless Communication Network", U.S. Patent 16/698,748, Nov. 27, 2019.

- E. Pateromichelakis, C. Zhou, P. Keshavamurthy, and K. Samdanis, "End-to-End QoS Optimization for V2X Service Localization", in *2019 IEEE Global*

Communications Conference (GLOBECOM), 2019, pp. 1–6.

- P. Keshavamurthy, P. Spapis, D. Dahlhaus, and C. Zhou, "Session-Enabled Joint Radio Resource Selection for Co-Operative Automated Driving", in *2019 IEEE 89th Vehicular Technology Conference (VTC Spring)*, Apr. 2019, pp. 1–5.

- P. Keshavamurthy, E. Pateromichelakis, D. Dahlhaus, and C. Zhou, "Cloud-Enabled Radio Resource Management for Co-Operative Driving Vehicular Networks", in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2019, pp. 1–6.

- A. Kaloxylos, P. Keshavamurthy, P. Spapis, and C. Zhou, "Context Aware Control Schemes for the Performance Improvement of V2X Network Slices", *International Journal on Advances in Telecommunications*, vol. 11, no. 1 & 2, pp. 65–75, 2018.

- P. Keshavamurthy, R. P. Manjunath, P. Spapis, D. Dahlhaus, E. Pateromichelakis, and C. Zhou, "Graph-Based Dynamic Zone Configurations for Resource Management in V2V Communications", in *2018 IEEE 88th Vehicular Technology Conference (VTC Fall)*, Aug. 2018, pp. 1–5.

Bibliography

[1] T. Litman, "Autonomous Vehicle Implementation Predictions", Victoria Transport Policy Institute, Canada, Jun. 2021.

[2] SAE International, "Surface Vehicle Recommended Practice", no. J3016, Jun. 2018.

[3] US Homeland Security Digital Library, "Self-Driving Cars: Levels of Automation", Mar. 2017.

[4] R. Hult, G. R. Campos, E. Steinmetz, L. Hammarstrand, P. Falcone, and H. Wymeersch, "Coordination of Cooperative Autonomous Vehicles: Toward Safer and More Efficient Road Transportation", *IEEE Signal Processing Magazine*, vol. 33, no. 6, pp. 74–84, Nov. 2016.

[5] 5GAA, "A Visionary Roadmap for Advanced Driving Use Cases, Connectivity Technologies, and Radio Spectrum Needs", Sep. 2020.

[6] K. Liang, J. Mårtensson, and K. H. Johansson, "Heavy-Duty Vehicle Platoon Formation for Fuel Efficiency", *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 4, pp. 1051–1061, Apr. 2016.

[7] M. Boban, A. Kousaridas, K. Manolakis, J. Eichinger, and W. Xu, "Connected Roads of the Future: Use Cases, Requirements, and Design Considerations for Vehicle-to-Everything Communications", *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 110–123, Sep. 2018.

[8] T. Li, J. Wu, C. Chan, M. Liu, C. Zhu, W. Lu, and K. Hu, "A Cooperative Lane Change Model for Connected and Automated Vehicles", *IEEE Access*, vol. 8, pp. 54 940–54 951, 2020.

[9] Y. Zheng, B. Ran, X. Qu, J. Zhang, and Y. Lin, "Cooperative Lane Changing Strategies to Improve Traffic Operation and Safety Nearby Freeway Off-Ramps in a Connected and Automated Vehicles Environment", *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4605–4614, 2020.

[10] S. E. Lee, E. C. Olsen, and W. W. Wierwille, "A Comprehensive Examination of Naturalistic Lane-Changes", Virginia Tech Transportation Institute/National Highway Traffic Safety Administration, USA, 2004.

[11] Q. Wang and C. Phillips, "Cooperative Collision Avoidance for Multi-Vehicle Systems using Reinforcement Learning", in *2013 18th International Conference on Methods Models in Automation Robotics (MMAR)*, 2013, pp. 98–102.

[12] 3GPP, "Enhancement of 3GPP Support for V2X Scenarios", TS 22.186 V15.3.0, Jun. 2018.

[13] A. Kousaridas, D. Medina, S. Ayaz, and C. Zhou, "Recent Advances in 3GPP Networks for Vehicular Communications", in *2017 IEEE Conference on Standards for Communications and Networking (CSCN)*, Sep. 2017, pp. 91–97.

[14] 3GPP, "Study on NR Vehicle-to-Everything (V2X)", TR 38.885 V16.0.0, Mar. 2019.

[15] IEEE, "IEEE Standard for Information Technology - Telecommunications and Information Exchange Between Systems - Local and Metropolitan Area Networks - Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", *IEEE Std 802.11-2007 (Revision of IEEE Std 802.11-1999)*, pp. 1–1076, Jun. 2007.

[16] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures (Release 14)", TS 36.213 V14.4.0, Sep. 2017.

[17] P. Rost, I. Berberana, A. Maeder, *et al.*, "Benefits and Challenges of Virtualization in 5G Radio Access Networks", *IEEE Communications Magazine*, vol. 53, no. 12, pp. 75–82, 2015.

[18] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On Multi-Access Edge Computing: A Survey of the Emerging 5G Network Edge Cloud Architecture and Orchestration", *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[19] G. Premsankar, M. Di Francesco, and T. Taleb, "Edge Computing for the Internet of Things: A Case Study", *IEEE Internet of Things Journal*, vol. 5, no. 2, pp. 1275–1284, Apr. 2018.

[20] R. Zhang, X. Cheng, Q. Yao, C. X. Wang, Y. Yang, and B. Jiao, "Interference Graph-Based Resource-Sharing Schemes for Vehicular Networks", *IEEE Transactions on Vehicular Technology*, vol. 62, no. 8, pp. 4028–4039, Oct. 2013.

[21] R. Fritzsche and A. Festag, "Location-Based Scheduling for Cellular V2V Systems in Highway Scenarios", in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, Jun. 2018, pp. 1–5.

[22] L. Liang, G. Y. Li, and W. Xu, "Resource Allocation for D2D-Enabled Vehicular Communications", *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3186–3197, Jul. 2017.

[23] W. Sun, E. G. Ström, F. Brännström, K. C. Sou, and Y. Sui, "Radio Resource Management for D2D-Based V2V Communication", *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 6636–6650, Aug. 2016.

[24] W. Sun, D. Yuan, E. G. Ström, and F. Brännström, "Cluster-Based Radio Resource Management for D2D-Supported Safety-Critical V2X Communications", *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2756–2769, Apr. 2016.

[25] M. Botsov, M. Klügel, W. Kellerer, and P. Fertl, "Location Dependent Resource Allocation for Mobile Device-to-Device Communications", in *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2014, pp. 1679–1684.

[26] H. Yang, L. Zhao, L. Lei, and K. Zheng, "A Two-Stage Allocation Scheme for Delay-Sensitive Services in Dense Vehicular Networks", in *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017, pp. 1358–1363.

[27] C. Wei, A. C. .-. Huang, C. Chen, and J. Chen, "QoS-Aware Hybrid Scheduling for Geographical Zone-Based Resource Allocation in Cellular Vehicle-to-Vehicle Communications", *IEEE Communications Letters*, vol. 22, no. 3, pp. 610–613, Mar. 2018.

[28] S. K. Gupta, J. Y. Khan, and D. T. Ngo, "Cluster-based D2D Architecture for Safety Services in Vehicular Ad hoc Networks", in *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Apr. 2018, pp. 43–48.

[29] M. I. Ashraf, M. Bennis, C. Perfecto, and W. Saad, "Dynamic Proximity-Aware Resource Allocation in Vehicle-to-Vehicle (V2V) Communications", in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6.

[30] M. I. Ashraf, C.-F. Liu, M. Bennis, and W. Saad, "Towards Low-Latency and Ultra-Reliable Vehicle-to-Vehicle Communication", in *2017 European Conference on Networks and Communications (EuCNC)*, Jun. 2017, pp. 1–5.

[31] X. Hou, Y. Li, M. Chen, D. Wu, D. Jin, and S. Chen, "Vehicular Fog Computing: A Viewpoint of Vehicles as the Infrastructures", *IEEE Transactions on Vehicular Technology*, vol. 65, no. 6, pp. 3860–3873, Jun. 2016.

[32] J. Feng, Z. Liu, C. Wu, and Y. Ji, "AVE: Autonomous Vehicular Edge Computing Framework with ACO-based Scheduling", *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 10 660–10 675, Dec. 2017.

[33] Y. Xiao and C. Zhu, "Vehicular Fog Computing: Vision and Challenges", in *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Mar. 2017, pp. 6–9.

[34] ETSI, "Multi-access Edge Computing(MEC); Study on MEC Support for V2X Use Cases", GR MEC 022 V2.1.1, Sep. 2018.

[35] 5GAA, "Toward Fully Connected Vehicles: Edge Computing for Advanced Automotive Communications", Dec. 2017.

[36] Z. Ning, X. Wang, and J. Huang, "Mobile Edge Computing-Enabled 5G Vehicular Networks: Toward the Integration of Communication and Computing", *IEEE Vehicular Technology Magazine*, vol. 14, no. 1, pp. 54–61, Mar. 2019.

[37] A. Aissioui, A. Ksentini, A. M. Gueroui, and T. Taleb, "On Enabling 5G Automotive Systems Using Follow Me Edge-Cloud Concept", *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5302–5316, Jun. 2018.

[38] J. Feng, Z. Liu, C. Wu, and Y. Ji, "HVC: A Hybrid Cloud Computing Framework in Vehicular Environments", in *2017 5th IEEE International Conference*

on Mobile Cloud Computing, Services, and Engineering (MobileCloud), Apr. 2017, pp. 9–16.

[39] J. Feng, Z. Liu, C. Wu, and Y. Ji, "Mobile Edge Computing for the Internet of Vehicles: Offloading Framework and Job Scheduling", *IEEE Vehicular Technology Magazine*, vol. 14, no. 1, pp. 28–36, Mar. 2019.

[40] K. Zhang, Y. Mao, S. Leng, Y. He, and Y. Zhang, "Mobile-Edge Computing for Vehicular Networks: A Promising Network Paradigm with Predictive Off-Loading", *IEEE Vehicular Technology Magazine*, vol. 12, no. 2, pp. 36–44, Jun. 2017.

[41] 5GAA, "Cellular V2X Conclusions based on Evaluation of Available Architectural Options", Dec. 2018.

[42] P. Keshavamurthy, E. Pateromichelakis, D. Dahlhaus, and C. Zhou, "Edge Cloud-Enabled Radio Resource Management for Co-Operative Automated Driving", *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 7, pp. 1515–1530, 2020.

[43] P. Keshavamurthy, E. Pateromichelakis, D. Dahlhaus, and C. Zhou, "Resource Scheduling for V2V Communications in Co-Operative Automated Driving", in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[44] S. Ayaz, D. Medina, P. Keshavamurthy, and C. Zhou, "Devices and Methods for Cloud-based Sidelink Scheduling and Base Station Interface Therefor", U.S. Patent 16/700,683, Dec. 2, 2019.

[45] S. Ayaz, D. Medina, P. Keshavamurthy, and M. Dillinger, "Devices and Methods for Communication in a Wireless Communication Network", U.S. Patent 16/698,748, Nov. 27, 2019.

[46] P. Keshavamurthy, E. Pateromichelakis, D. Dahlhaus, and C. Zhou, "Cloud-Enabled Radio Resource Management for Co-Operative Driving Vehicular Networks", in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Apr. 2019, pp. 1–6.

[47] P. Keshavamurthy, P. Spapis, D. Dahlhaus, and C. Zhou, "Session-Enabled Joint Radio Resource Selection for Co-Operative Automated Driving", in *2019 IEEE 89th Vehicular Technology Conference (VTC Spring)*, Apr. 2019, pp. 1–5.

[48] P. Spapis, P. Keshavamurthy, and C. Zhou, "User Equipment for V2V Communications", pat. PCT/EP2018/077301, Apr. 16, 2020.

[49] P. Keshavamurthy, R. P. Manjunath, P. Spapis, D. Dahlhaus, E. Pateromichelakis, and C. Zhou, "Graph-Based Dynamic Zone Configurations for Resource Management in V2V Communications", in *2018 IEEE 88th Vehicular Technology Conference (VTC Fall)*, Aug. 2018, pp. 1–5.

[50] A. Kaloxylos, P. Keshavamurthy, P. Spapis, and C. Zhou, "Context Aware Control Schemes for the Performance Improvement of V2X Network Slices", *International Journal on Advances in Telecommunications*, vol. 11, no. 1 & 2, pp. 65–75, 2018.

[51] E. Pateromichelakis, C. Zhou, P. Keshavamurthy, and K. Samdanis, "End-to-End QoS Optimization for V2X Service Localization", in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[52] C. Mohanram and S. Bhashyam, "Joint Subcarrier and Power Allocation in Channel-Aware Queue-Aware Scheduling for Multiuser OFDM", *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, pp. 3208–3213, Sep. 2007.

[53] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing Quality of Service over a Shared Wireless Link", *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, 2001.

[54] Q. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, "On Utility Models for Access Network Selection in Wireless Heterogeneous Networks", in *NOMS 2008 - 2008 IEEE Network Operations and Management Symposium*, Apr. 2008, pp. 144–151.

[55] 3GPP, "Application Layer Support for V2X Services; Functional Architecture and Information Flows", TS 23.286 V1.0.0, Mar. 2019.

[56] 3GPP, "Service Enabler Architecture Layer for Verticals; Functional Architecture and Information Flows", TS 23.434 V1.0.0, Mar. 2019.

[57] F. P. Kelly, A. K. Maulloo, and D. K. Tan, "Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability", *Journal of the Operational Research society*, vol. 49, no. 3, pp. 237–252, 1998.

[58] J. W. Lee, M. Chiang, and A. R. Calderbank, "Price-based Distributed Algorithms for Rate-Reliability Tradeoff in Network Utility Maximization", *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 5, pp. 962–976, May 2006.

[59] Y. Li, M. Chiang, A. R. Calderbank, and S. N. Diggavi, "Optimal Rate-Reliability-Delay Tradeoff in Networks with Composite Links", in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007, pp. 526–534.

[60] S. Kandukuri and S. Boyd, "Optimal Power Control in Interference-Limited Fading Wireless Channels with Outage-Probability Specifications", *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 46–55, Jan. 2002.

[61] L. Liu, S. Lu, R. Zhong, B. Wu, Y. Yao, Q. Zhang, and W. Shi, "Computing Systems for Autonomous Driving: State of the Art and Challenges", *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6469–6486, 2020.

[62] D. Nace and M. Pioro, "Max-Min Fairness and its Applications to Routing and Load-Balancing in Communication Networks: A Tutorial", *IEEE Communications Surveys Tutorials*, vol. 10, no. 4, pp. 5–17, 2008.

[63] M. Grötschel and Y. Wakabayashi, "Facets of the Clique Partitioning Polytope", *Mathematical Programming*, vol. 47, no. 1-3, pp. 367–387, 1990.

[64] A. Mehrotra and M. A. Trick, "Cliques and Clustering: A Combinatorial Approach", *Operations Research Letters*, vol. 22, no. 1, pp. 1–12, 1998.

[65] E. Pateromichelakis, M. Shariat, A. Quddus, M. Dianati, and R. Tafazolli, "Dynamic Clustering Framework for Multi-Cell Scheduling in Dense Small Cell Networks", *IEEE Communications Letters*, vol. 17, no. 9, pp. 1802–1805, Sep. 2013.

[66] D. Krajzewicz, J. Erdmann, M. Behrisch, and L. Bieker, "Recent Development and Applications of SUMO - Simulation of Urban MObility", *International Journal On Advances in Systems and Measurements*, vol. 5, no. 3 & 4, pp. 128–138, Dec. 2012.

[67] 3GPP, "Study on Evaluation Methodology of New Vehicle-to-Everything V2X Use Cases for LTE and NR", TR 37.885 V15.1.0, Sep. 2018.

[68] 3GPP, "Study on LTE-based V2X Services", TR 36.885 V14.0.0, Jun. 2016.

[69] I. Bello, H. Pham, Q. V. Le, M. Norouzi, and S. Bengio, "Neural Combinatorial Optimization with Reinforcement Learning", 2017.

[70] N. Mazyavkina, S. Sviridov, S. Ivanov, and E. Burnaev, "Reinforcement Learning for Combinatorial Optimization: A Survey", *Computers & Operations Research*, vol. 134, 2021.