



Article

5G V2X System-Level Architecture of 5GCAR Project

Massimo Condoluci ^{1,*} , Laurent Gallo ², Laurent Mussot ², Apostolos Kousaridas ³,
Panagiotis Spapis ³, Maliheh Mahlouji ⁴ and Toktam Mahmoodi ⁴

¹ Ericsson Research, Torshamnsgatan 23, 164 83 Stockholm, Sweden

² Orange Labs Networks, Orange Gardens, 44 avenue de la République, 92326 Châtillon, France; laurent.gallo@orange.com (L.G.); laurent.mussot@orange.com (L.M.)

³ Huawei Technologies, Riesstraße 25, 80992 Munich, Germany; apostolos.kousaridas@huawei.com (A.K.); panagiotis.spapis@huawei.com (P.S.)

⁴ Department of Engineering, King's College London, Strand, London WC2R 2LS, UK; maliheh.mahlouji@kcl.ac.uk (M.M.); toktam.mahmoodi@kcl.ac.uk (T.M.)

* Correspondence: massimo.condoluci@ericsson.com

Received: 8 October 2019; Accepted: 16 October 2019; Published: 19 October 2019



Abstract: One of the goals of the 5G Communication Automotive Research and innovation (5GCAR) project has been to evaluate and propose system architecture enhancements aiming at supporting the strict requirements of vehicle-to-everything (V2X) use cases. In this paper, we provide an overview of 3GPP 5G system architecture, which is used as a baseline architecture in the project, and we present the main architectural enhancements introduced by 5GCAR. The work of the project focused on the following categories: (i) end-to-end security, also including aspects of privacy; (ii) network orchestration and management; (iii) network procedures; (iv) edge computing enhancements; and (v) multi-connectivity cooperation. The enhancements introduced by 5GCAR to above-listed categories are discussed in this paper, while a more detailed analysis of some selected features is presented.

Keywords: V2X; 5G; QoS; system architecture

1. Introduction

The automotive sector is considered to be one of the most prominent verticals that will benefit from the capabilities of the upcoming 5G cellular networks [1,2]. Vehicular applications cover a wide range of use cases and thus a large set of associated requirements. Examples include very high data rates and timely service delivery, while also considering ultra-low communication latencies, just to mention a few. Complex scenarios where vehicles communicate among themselves and also with nearby road infrastructure, road users, clouds, etc.—also known as vehicle-to-everything (V2X) communications—will not only leverage 5G network but will play a key role in its design. The H2020 5G PPP Phase 2 project 5G Communication Automotive Research and innovation (5GCAR) [3] worked towards the definition of enhancements in terms of system architecture, security, and privacy, specifically targeting automotive applications. In particular, 5GCAR considered five different classes of use cases: cooperative maneuver, cooperative perception, cooperative safety, autonomous navigation, and remote driving [4]. Among the possible use cases of these classes, 5GCAR considers five use cases (lane merge, see-through, network-assisted vulnerable pedestrian protection, high definition local map acquisition, and remote driving for automated parking) to derive the target requirements to be used to drive the network design as well as the final demonstration of the project outcomes [5].

One of the goals of 5GCAR is to study and propose target enhancements of the 3GPP 5G service-based architecture (SBA) [6] in order to support the large set of requirements of V2X applications. The project analysed the technical and non-technical issues brought by V2X requirements that challenge

the current architectural conception. The general technical challenges brought by vehicular applications were summarized in four main technical areas: (i) V2X main characteristics (mobility and simultaneous requirements for massive numbers of ultra-reliable, low latency, high bandwidth communications); (ii) multiple access network connectivity (including the management of multiple network slices); (iii) resilience requirements, security, and data privacy; and (iv) roaming between operators. 5GCAR proposed a collection of enhancements, referred to as “technical components”, designed to address specific needs and requirements of automotive applications. The introduced enhancements focus on optimized end-to-end V2X network connectivity for highly reliable and low-latency V2X services, security and privacy, (Quality of Service) QoS and traffic flow management in a multi-Radio Access Technology (RAT) and multi-link V2X communication system [7]. The contribution from 5GCAR project has been to address enhancements that can be mapped to 3GPP system architecture, i.e., proposed technical components can be considered as add-ons to be integrated into already defined 3GPP network functions. Above mentioned technical components were grouped into five different categories:

- end-to-end security;
- network orchestration and management;
- network procedures;
- edge computing enhancements;
- multi-connectivity cooperation.

The aim of this paper is to present the outputs of the studies performed by 5GCAR on the system architecture.

The remainder of the paper is structured as follows. Section 2 summarizes the 3GPP 5G system architecture, which has been used as a baseline within 5GCAR. Section 3 provides an overview of the overall 5GCAR architecture design, in which Section 4 offers a more detailed description of some selected components. Finally, Section 5 provides conclusive remarks.

2. GPP 5G System Architecture

The 5G system architecture, shown in Figures 1 and 2, has been designed by 3GPP [6,7] by means of defining network functions (NFs) instead of network entities, with the further novelty that the control plane (CP) is centered around services instead of interfaces. The latter feature allows each function to offer web-based services, utilizing representational state transfer (REST) APIs, and other functions can access such services through subscription procedures. In the remainder of this Section, we provide an overview of the functionalities of the NFs of the 3GPP 5G system architecture which has been assumed as a baseline for the architectural work in 5GCAR.

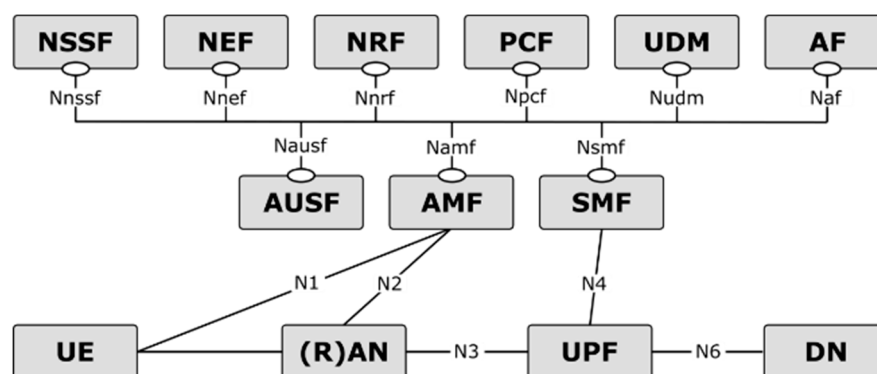


Figure 1. 5G system architecture [6].

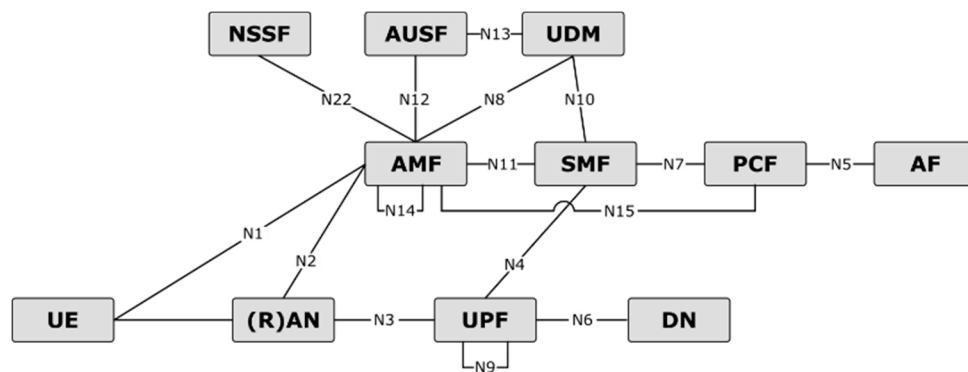


Figure 2. Reference point representation of the 5G system architecture [6].

The network functions repository function (NRF) is used in order to collect and maintain information of the NF instances which are available at any time in the network. Examples of such information are NF profiles and addresses. The NRF supports NF service discovery.

The policy control function (PCF) provides a unified policy framework governing the network behaviour, with tasks such as providing the other CP functions with the policy rules to be enforced. Furthermore, the PCF acts as the front-end interface for other NFs to access subscription information relevant for policy decisions which are stored in the unified data repository (UDR).

The session management function (SMF) handles all of the functionalities related to the management of Protocol Data Unit (PDU) Sessions, such as establishment, modification, and release. Additional functionalities include maintaining tunnels between the (R)AN and user plane (UP) nodes, IP addresses allocation, and management.

The access and mobility management function (AMF) is the termination point of the non-access stratum (NAS), performing functionalities such as User Equipment (UE) registration management, mobility management, reachability, connection management, user authentication, and access authentication [7]. Furthermore, the AMF handles session management messages coming from the core network towards the RAN.

The unified data management (UDM) manages information related to the UE and to the subscriber profile. Among the tasks of the UDM, there are the 3GPP Authentication and Key Agreement credentials, user identification handling, and storing of data related to registration management of NFs serving the UE.

The network data analytics function (NWDAF), provides other NFs with analytics information on the network behaviour, such as load level information on a slice level.

The authentication server function (AUSF) implements authentication and security functions.

The network slice selection function (NSSF) is an NF introduced in 5G to determine the set of network slices that a certain UE is allowed to access. In addition, the NSSF determines the set of candidate AMFs suitable to serve the UE.

The application function (AF) is a CP function interacting with other NFs of the 5G core (5GC) network. The AF implements procedures such as influence of traffic routing of PDU sessions based on application-specific requirements. According to the deployment configuration, AFs can be trusted or untrusted. In the former case, they may be located within the operator's network and directly connected to the other control plane functions. The latter deployment option is for AFs deployed in a public network. In this case, the AF interacts with the 5GC via the network exposure function (NEF), which is responsible for exposing network functionalities and capabilities and handles masking of sensitive internal user information before information exposure.

The user plane function (UPF) is the only node in UP and implements functionalities of packet routing and forwarding. It is responsible for the enforcement of UP policy rules, and in general QoS enforcement for the UP. The UPF acts as an anchor point for intra/inter-RAT mobility, and an external point of interconnect to a data network for PDU sessions.

More details on the functionalities of the NFs of the 5G system architecture can be found in [6], while more details on the procedures of the NFs are listed in [7].

3. Overall 5GCAR Architecture Design

The work in 5GCAR with regards to the system architecture has focused on five different areas of enhancements. A visual representation of the technical components introduced by 5GCAR is given in Figure 3, which are split considering their relevant area of enhancements. The enhancements introduced for each area, with more details of the associated features introduced by 5GCAR, are discussed in the remainder of this section. For more details on the system architecture enhancements introduced by 5GCAR, please refer to [8], which also provides analyses of pros and cons of the technical components, more information about open challenges related to their implementation and integration with existing 3GPP network functions.

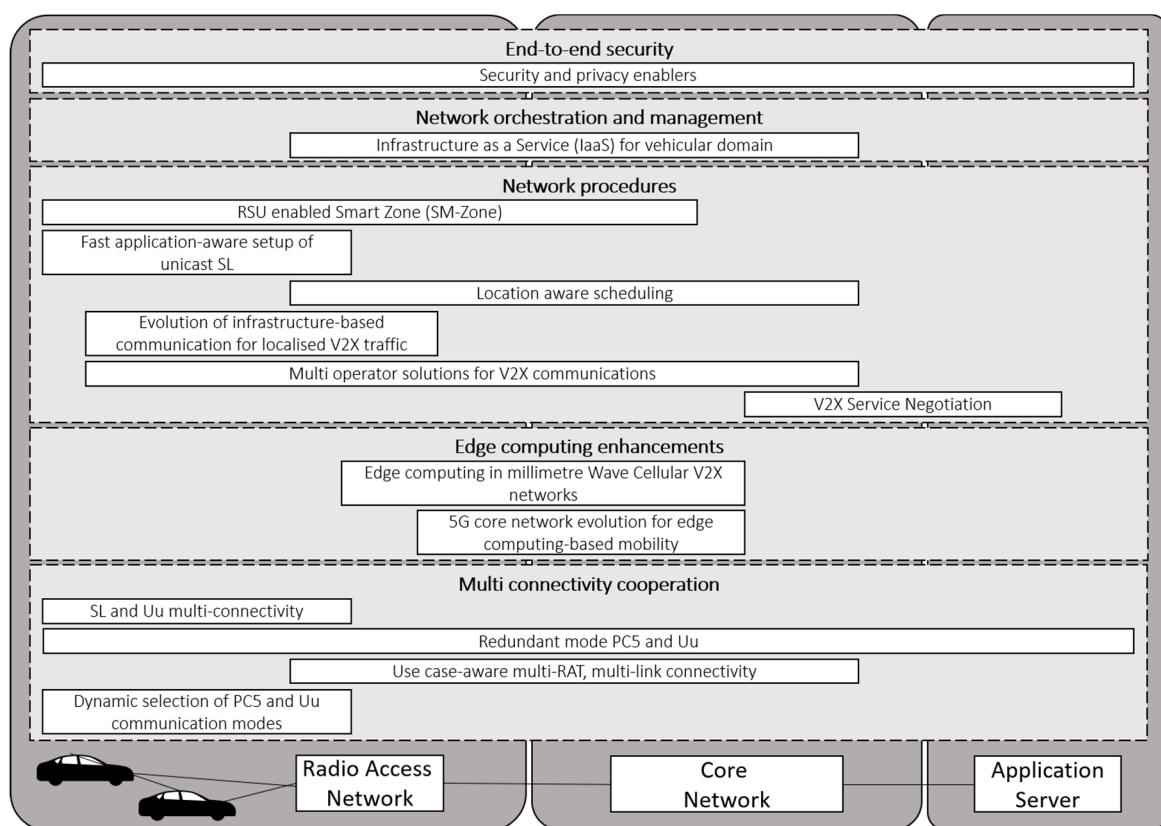


Figure 3. The 5G Communication Automotive Research and innovation (5GCAR) enhancements for the end-to-end system architecture, highlighting the areas the different enhancements refer to.

3.1. End-To-End Security

Two key requirements for V2X communication and applications are security and privacy, considering also the impact that procedures and mechanisms adopted to handle security and privacy might have on the communications. For instance, concerns might be associated with some security operations that may introduce additional latency or introduce further traffic in the communication. An example of such possible implications is signing all messages with a hardware security module (or USIM—universal subscriber identity module). Signing or verifying a signature, encrypting or decrypting a message also requires CPU-intensive cryptographic operations that might negatively impact latency, even if cryptographic hardware accelerators and hardware security modules (HSMs) are used. Another example is authentication, authorization, and accounting that might require a

signature and its associated certificate to be attached to some (or all) messages. This may increase the length of a message of a few hundred bytes with a consequent overhead impact.

To mitigate the impacts discussed above, session-based communications introduce benefits, for instance, authentication and authorization are performed once at the beginning of the session. Then session keys are exchanged and/or derived to encrypt remaining exchanges. Nevertheless, this approach works fine for vehicle-to-network (V2N, i.e., UE to V2X server application), or one-to-one V2V/I2V/V2I applications but it is not suitable for one-to-many V2V or I2V/V2I applications (for instance, for exchanging awareness messages). To overcome this limitation, one approach, which is used in Intelligent Transportation System (ITS) for (Common Awareness Messages) CAM, DENM, MAP, SPaT, or IVI messages over ITS-G5, is to sign every single message. In this case, two drawbacks associated with the limitations discussed above should be highlighted: (i) Signing every message with an HSM has a latency penalty, and (ii) every receiving UE needs to go through all consistency and relevance verification checks introducing further processing delay. A second approach is to rely on a central entity in charge of performing authentication and authorization checks, and to provide a session key for the group of communicating UEs, as for instance considered in 3GPP for ProSe one-to-many security communication [9].

On this topic of one-to-many end-to-end security, 5GCAR evaluated the feasibility for V2X use cases of the usage of (strong) authentication and authorization performed against a central entity, where the security is handled on a service-basis (i.e., each V2X application wanting to join a certain service will need a group of keys devoted to the specific application considered). The main principles are now discussed.

A central entity acting as a key manager and managing the security for a certain application/service is introduced. We refer to this central entity as the “Key Manager”. The approach is agnostic of who is responsible for this entity (MNOs, public administrations, OEMs, OTT-like entities, etc.). Multiple Key Managers may be deployed, each one operating in a given geographical area. The connection to the central entity is secured by using the UE pseudonym certificates which could be provided in the same way as it is done in the current ITS solution. The requirements are a secure process for the UE certificate enrolment and the trust of the authority chain by the central entity. The central entity may be located in the back office, or as part of road infrastructure, or replaced by one vehicle that plays a specific role in the scope of one specific V2X application (such as the lead vehicle in a platoon).

Figure 4a depicts how the Key Manager generates a high number of keys, called Z , which are then used as shared secret keys by all vehicles wishing to participate in a given V2X service. For security purposes, such Z keys have a short validity period. The validity of Z keys may be inversely proportional to the number of vehicles in the group that shares the same key Z (Further discussion and service adaptation will be needed to define Z keys validity duration, and how far in the future a vehicle can request and download Z keys batches. In addition, other aspects such as integrity need to be further evaluated). Trust is based on the knowledge of such shared secret keys during its validity period. To compute the shared keys, an elliptic curve is used defined by a base point G and a prime p , z is a random value, Z keys are curve points computed from a random number z : $Z = G^z \bmod p$. The three elements (G , p , and z) are specific to the V2X service and the backend entity. The vehicles need to regularly connect to the Key Manager to download batches of signed tuples containing the curve specifiers G and p , a key Z and its associated validity period. The interaction between the vehicle and the Key Manager (i.e., request and download of batches) is performed over a secured connection, typically TLS based: the vehicle and Key Manager authenticate each other using their respective ETSI V2X certificates (pseudonym certificate for the vehicle). The Key Manager may also verify whether the UE pseudonym certificate holds permissions to request a batch for a given V2X service.

Figure 4b depicts how the shared secret key Z is used during its validity period. The considered example is when one vehicle sends a message to all other vehicles participating to the V2X service for which key Z has been obtained. Key Z is used to derive another key K_z that may be used to (i) encrypt the message payload and compute a hash for integrity purposes and to (ii) encrypt only the hash for

integrity purposes. The vehicle generates a random number “a” which is used to derive the encryption key Kaz. For non-repudiation purposes, UE may calculate $A = (g)a \pmod{p}$ and upload the result to any entity responsible for traceability.

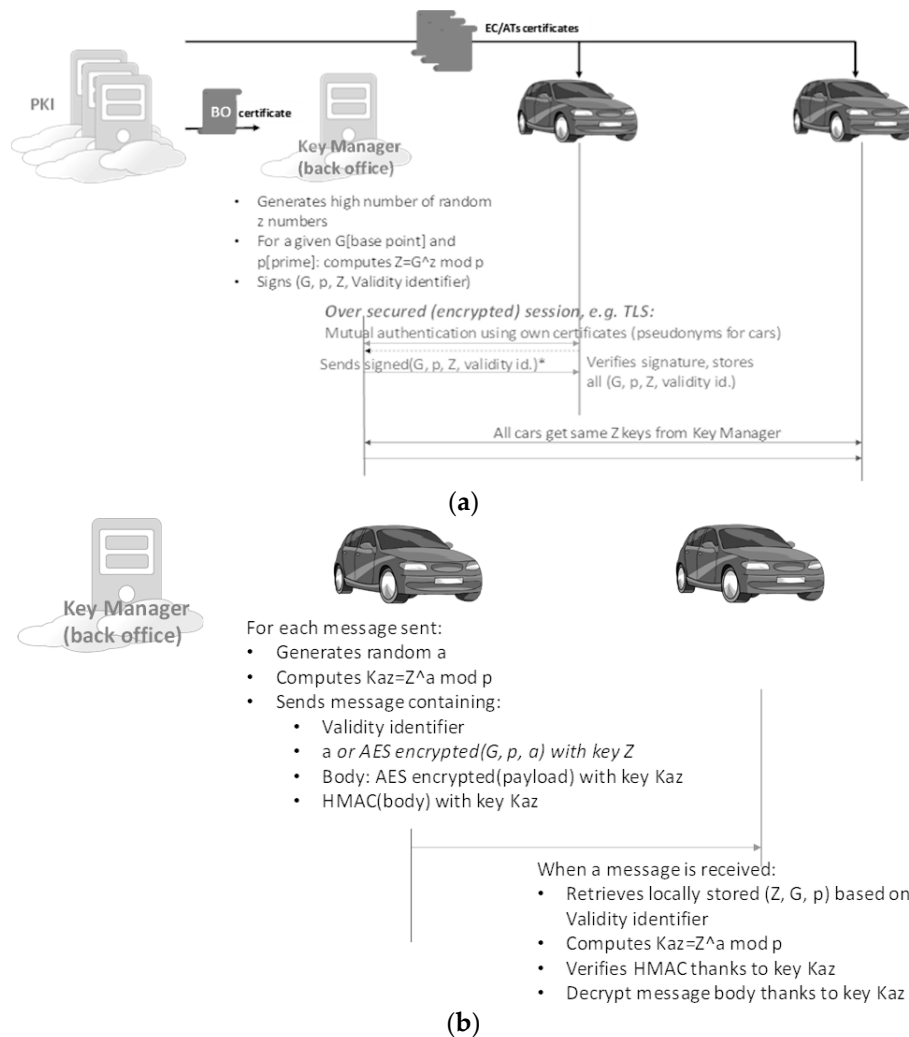


Figure 4. (a) Vehicle gets shared keys from the back-office Key Manager and (b) anonymous messages exchange between vehicles.

The benefits introduced by this approach are in terms of increased efficiency from a bandwidth and latency point of view. Furthermore, the proposal also aims to reduce the cost associated with security by not requiring as many pseudonym certificates per UE as currently envisioned. Indeed, based on current E.U. policies, up to 100 pseudonym certificates may be required for every vehicle, every week (each pseudonym certificate has a validity period of at most one week). The cost of generating, and especially downloading to the vehicle this large number of certificates represent a significant cost for an OEM when considering millions of vehicles being deployed.

Another topic investigated by the 5GCAR project has been privacy in the context of V2X. The European Commission draft Delegated Act emphasizes that C-ITS services must comply with the E.U. law on the protection of personal data, in particular the General Data Protection Regulation (GDPR). When applied to V2X, personal data should be, not only intended as information allowing to identify a person (e.g., driver, passenger, or a vehicle’s owner), but also information such as geographical location data, mileage, driving style, or other vehicle’s technical data. In this context, aspects of which data are requested, whether they are stored (and if they are, for how long) need careful consideration. Data shall not be stored unless required to provide a service (data minimization).

V2X protocols and messages that have been standardized so far support the adoption of pseudonyms. The utilisation of pseudonyms and anonymization are not quite the same: for misbehaviour reporting purposes, there must exist some entity that can link a pseudonym to its enrolment certificate (i.e., true identity). It must be noted that the draft Delegated Act does not address how V2X devices, protocols, and applications should comply with GDPR, it only states that they must comply with GDPR. As a consequence, vehicle manufacturers may have to provide the ability for a driver to disable V2X. GDPR applicability to V2X should therefore be clarified. As one example, CAMs will be critical for many day-2 safety applications, but it is not clear whether they comply with GDPR. For the sake of safety-related applications, regulation should allow a minimum set of information to be always sent by devices installed in vehicles or other road user equipment, in order to guarantee that a minimum set of safety-critical services can be operated.

3.2. Network Orchestration and Management

Network orchestration and management are two important aspects to be optimized for network deployment and management, which obviously impact achievable network performance. The network infrastructure will be composed of servers distributed at several locations ranging from the edge (considering local/edge cloud capabilities) to district level (big town, or group of towns), to regional level, country level, and up to more centralized locations (the Internet for instance). This raises the challenge of defining the whole system in a coherent way. The orchestration of such infrastructure should be able to define a set of infrastructure services enabling to automate deployments, maintenance, management, configuration, upgrade, and repair of each component of the system.

The baseline considered by the 5GCAR project with respect to orchestration is composed of the different open source NFV/SDN frameworks such as:

- ETSI open source MANO (OSM), which is an open-source management and orchestration (MANO) framework with a strong relationship to industry. The focus is on the provision of network services on top of OpenStack/VMware/AmazonWS infrastructure. Its current release (release 4) is production-ready, but it lacks support for containers and service function chaining (SFC).
- Open network automation platform (ONAP), which is defined as an open-source software platform that delivers capabilities for the design, creation, orchestration, monitoring, and life cycle management of virtual network functions. The carrier-scale software-defined networks that contain them and higher-level services that combine the above. ONAP has consequently a wider perimeter than ETSI OSM. It is expected that release 3 (Casablanca) would be fully functional, but current deployment requires a large number of resources.

Another aspect considered in the 5GCAR project is that the ITS software must be designed in accordance with cloud ready principles [10]. This adaptation to the features of the virtual resources and to the cloud tools allows reaching the requirements of the ITS software, notably the resiliency requirement.

The problem of the placement of components in a distributed environment is complex. An orchestrator will be used to deploy the components in order to satisfy running time SLA constraints such as latency, reliability, etc. Some open-source projects such as OpenStack Tricircle [11] deal with the problem of placement in a distributed environment. The works [12,13] are addressing the problem of placement of software components in a distributed environment.

Another aspect to be considered related to network management is the relationship with the variation of road traffic conditions. For example, a road may experience some ten-folds increase of vehicle density compared to that of preferred road traffic situations (for which placement of functionalities and resources has been planned). This brings us back to the need of having a network able to dynamically reconfigure itself to adapt to varying load conditions.

On this topic, 5GCAR analyzed how network orchestration and management capabilities can be enhanced in order to cope with the unique requirements of vehicular use cases. The main target has been to improve network management and reconfigurability to cope with the dynamicity in

terms of traffic demand of vehicular scenarios. On this topic, the focus of the 5GCAR project has been on aligning the orchestration framework with data analytics and exposure frameworks of the 3GPP system.

A high-level procedure for aligning orchestration with 3GPP data analytics and exposure frameworks can be found in Figure 5. At the application level, an interface will grant exchange of information between the application function (AF) and 3GPP functional control plane entities through the exposure framework, i.e., the NEF. In this case, the orchestration framework is considered as an AF interacting with the 3GPP exposure framework. For instance, examples of information that the NEF can expose to be used for network orchestration might include channel quality info, UE Location, Cell-ID, cell load, average network latency at access, core level, and target Cell-ID (before handover execution). From this point of view, the knowledge in advance of the target Cell-ID will be useful for instance at AF to correlate car trajectory with “network trajectory”, to be able to anticipate the fading situation in particular area (tunnel, building shadowing). The AF could be also able to organise in advance switching from one edge server to another if the target cell is attached to a different edge server.

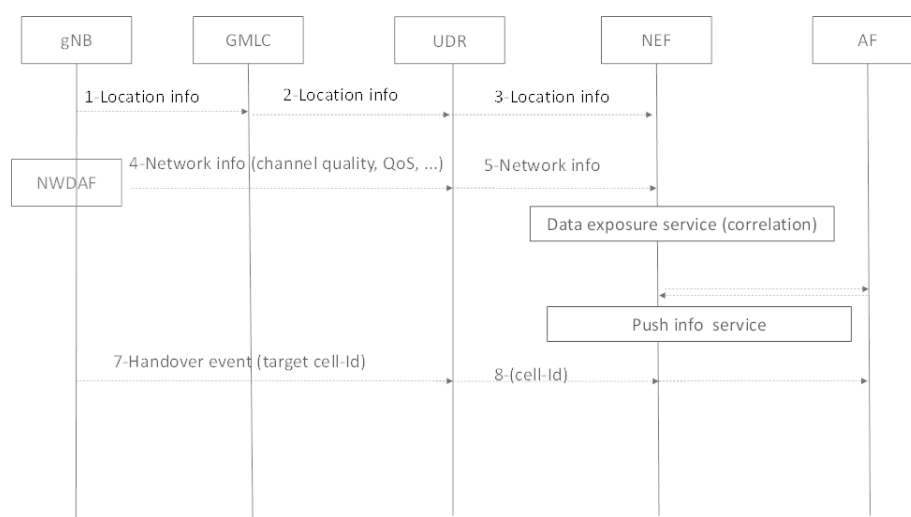


Figure 5. Impact over 3GPP interfaces and functional entities of aligning orchestration with data analytics and exposure framework.

The information exposed by the NEF can be collected from several sources. For instance, the next generation NodeB (gNB) could send information (radio layer info) towards the UDR, which could be then exposed by the NEF. The gateway mobile location centre (GMLC) defined in 3GPP Release 15 will also push location information towards the UDR. Some network information to be exposed could be generated by NWDAF, which could provide slice-granularity information in its current 3GPP specification.

Furthermore, APIs could be set between NEF and business support system (BSS) and/or operating support system (OSS), which will be in charge of pushing information towards the service orchestration framework in order to manage network components in an optimal way. In a similar way, APIs between AF and 3rd party applications can be used in order to push information towards 3rd party application orchestration framework.

One example of a use case where we can apply the above-mentioned alignment of orchestration capabilities with 3GPP exposure framework is the lane merge scenario. In this case, a specific traffic orchestrator running in a regional data centre can be instantiated. If the virtual infrastructure manager (VIM) [14] detects that the resource allocation for the orchestrator does not allow the ability to meet the service latency requirement (time for message processing is too high for instance) due to server high occupancy, it can then decide to redeploy the application component of the traffic orchestrator

to an edge cloud location also considering network performance information. Such a scenario can be obtained, for instance, by considering information exposed by the mobile network given by NEF, for instance, in terms of achievable QoS such as estimated or predicted latency, bitrate, etc. In this scenario, the service orchestration can take into consideration information about estimated QoS exposed by the NEF, and jointly use such information with other information from the VIM to estimate in advance whether, for instance, end-to-end latency for a given service topology (service component placement) can be met. In this scenario, the service orchestrator will then be able to select the best edge server to perform the vehicle traffic orchestration for a given set of cross-roads in accordance with the service level requirements of lane merge, also considering network performance measurements and/or information in addition to cloud-based information.

The network orchestration can be further enhanced by considering latency and reliability driven service orchestration. A similar approach has been used in [15], where to improve both the reliability and latency performance, a new network slicing solution that extends from resource slicing to service and function slicing is presented for 5G autonomous vehicular networks.

3.3. Network Procedures

Service delivery in mobile networks involves several procedures, such as network-application negotiation, connectivity establishment, mobility and user plane management, etc. [7]. Such procedures directly impact the service delivery, for instance in terms of introduced latency, while other impacts might be associated with required improvements to better handle V2X services. From this point of view, 5GCAR introduced several features in terms of network procedures, introducing the following enhancements: (i) improved management of roaming scenarios; (ii) improved awareness of service requirements and additional related information for the network when delivering a certain vehicular use case; (iii) optimization of control and user plane paths; (iv) optimization of scheduling capabilities considering vehicle mobility; (v) providing resource configuration stability for local services; and (vi) network-driven setup of unicast sidelink. The technical components associated with enhancements (i–iv) listed above will be discussed in more detail in Section 4. In particular, the enhancements introduced by 5GCAR on the interaction between network and applications will be given in Section 4.1, while additional improvements achievable by using enhanced application and service information are provided in Sections 4.2 and 4.3. More details on enhanced procedure allowing the ability to shorten the UP path for localized V2X communications are given in Section 4.4. Finally, more details on the enhancements for multi-operator scenarios will be provided in Section 4.5. The remainder of this Section will provide more information on enhancements (v) and (vi). One enhancement introduced by 5GCAR is related to the concept of SM-Zones as a solution to provide resource stability for local services that are provided by using roadside units (RSUs), either in the form of UE- or eNB-type RSUs. In this context, 5GCAR considers RSUs as logical architectural nodes delivering local services. The RSUs, as considered in 5GCAR, can be grouped to form smart radio access service areas, referred to as Smart Zones (SM-Zones). SM-Zones might be configured to be local to targeted individual roads (covering just road areas). In this case, radio transmissions, as well as the end-to-end service flows of V2X communications between UEs on roads, may be kept local to individual roads, which is rather desirable for the efficient utilisation of spectrum and network resources while meeting QoS requirements of advanced V2X services. This comes with the deployment cost of RSUs along targeted individual roads, to be mounted on roadside lamps for instance. While the exploitation of RSU may bring benefits in terms of increasing overall capacity of supported V2X traffic acting as “local hot spots” for local services, an aspect that might cause performance variation is related to different resource configurations associated with different RSUs. From this point of view, a vehicle may experience performance variation/degradation when moving from one RSU to another. The goal of resource stability among the RSUs within an SM-Zone is achieved by the enhancement that, when entering a certain SM-Zone, a UE is provided with a communication configuration which is kept within the whole SM-Zone even if the zone is composed of several RSUs. This allows the UE to achieve stable

communication within the whole SM-Zone. Given the local nature of SM-Zones, ad-hoc configurations tailored for specific SM-Zones can be considered to optimize performance and network resource utilization considering the local connectivity requirements and scenarios (expected vehicle density, road traffic pattern, service characteristics, etc.). The work done in 5GCAR with respect to SM-Zone optimization can be further extended considering approaches similar to those in [16], where aspects such as user pause probability, user arrival, and departure probabilities are derived and used for evaluating the user mobility performance in a hotspot-type 5G small cell network. Furthermore, the work in [16] derives the coverage probabilities of small cell and macro cell BSs for all users in 5G small cell networks, respectively. The approach in [16] can be tailored for SM-Zone and extended for resource configuration selection among multiple RSUs belonging to the same SM-Zone.

Another network procedure improvement introduced by 5GCAR is related to the facilitation of the establishment of unicast communication via sidelink (SL). Currently, unicast communication over SL may be realized by using L1 broadcast SL [9] and using ad-hoc solutions for obtaining unicast features on higher layer—e.g., non-access-stratum or application layer functionalities allowing a UE to discard a message if addressed to a different UE. In this option, the setup of the unicast communication session over SL relies also on non-access-stratum or application-layer signalling, including the discovery phase of the involved UEs, procedures which involve side effects in terms of time consumption thus reducing overall efficiency. In addition, due to using L1 broadcast SL, UEs non-involved in the unicast communication but which are in proximity of the two unicast UEs may need to process and then discard the unintended unicast communications. The utilization of L1 unicast SL, instead of L1 broadcast SL with unicast features at higher layers, may overcome the limitations discussed above under the assumption that each of involved UEs obtain a locally unique L1 ID, allowing that a Tx UE of the L1 unicast SL may schedule an L1 unicast SL transmission addressed to only a targeted Rx UE by indicating the locally unique L1 ID of the targeted Rx UE. Given that the assignment of such locally unique IDs needs to be guaranteed, in practice this relies on assistance from a serving network, given that other solutions such as using a globally unique UE ID such as international mobile subscriber identity (IMSI) or application ID as L1 ID involve high L1 signalling overhead as well as security implications. In this regard, 5GCAR evaluated the introduction of some network procedures to be included in RAN for facilitating and speeding up the setup of an SL unicast communication (i.e., network-driven SL unicast setup). In particular, the following was considered:

- The serving RAN may indicate cell-specific support for assisting the L1 unicast SL in the signalling of system information, signalling which may include allocation of L1 ID and SL resources for L1 unicast SL with constraints such as the maximum number of unicast SLs per an SL communication session.
- In case the initiator UE is in the network coverage, the UE decides whether to use L1 unicast SL and performs a request towards the serving RAN, a request which may include also a request for an SL resource allocation. The response from RAN may provide a resource allocation or, otherwise, an indication that requested network assistance, for the time being, cannot be provided.
- As a fallback solution, if the initiator UE is out of coverage of the RAN is not able to assist the setup of L1 unicast SL, L1 broadcast SL is used.

3.4. Edge Computing Enhancements

The availability of computing capabilities at the edge of the network together with an overall management of central and edge resources opens up the possibility for several improvements in mobile networks to support vehicular use cases [17]. To take full advantage of edge computing, 5GCAR considered several enhancements from a core network perspective as well as from an access network point of view. From a core network perspective, 5GCAR enhanced the concept of mobility support with an approach where mobility-related network procedures (i.e., handover) are synched with mobility-related “edge” procedures (i.e., job migration). Another enhancement introduced by 5GCAR is related to the optimization of task execution jointly considering network and edge resources.

The first enhancement is associated with mobility support. Being a highly mobile terminal by definition, a vehicle may require frequent handovers when moving at high speed on a highway, for instance. Given that an edge cloud is usually hosted on local and small size data centres designed to serve a relatively limited small number of vehicles attached to base stations in its proximity. This involves the side effect that vehicles may handover multiple times within a single journey, within access points connected to different edge cloud infrastructures. This might involve several limitations, as when migrating from one edge cloud to another also the application needs to be migrated accordingly and be updated with the appropriate internal state before the vehicle completes the handover phase. From this point, considering a deployment where edge clouds co-located at some edge UPFs are available, 5GCAR considered enhancing handover procedures in 3GPP by jointly considering also edge cloud re-location during mobility-related procedures. The UE is configured to constantly send some particular information to the RAN (location reporting, channel status reports, etc.) and such information is used to understand whether a handover should be triggered, in this case using, for instance, the location reporting to check whether an edge cloud closer to UE location is available (for instance, this could be achieved by using the data analytics and exposure framework of the 3GPP system). In this scenario, when there is a signal power decrease and UE handover towards a target RAN, the RAN sends the information that the UE has moved to a new target cell to the AMF, which then starts the process of choosing a target SMF. In this case, this selection triggers the process of choosing the appropriate target edge UPF with adequate edge cloud capabilities, this happening together with the reservation of resources and transfer of data between the source and target elements. With this enhancement, edge cloud re-location is intrinsically supported directly with UPF re-selection (i.e., join network and edge mobility support).

Another aspect considered by 5GCAR is the optimization of task execution jointly considering network and edge resources in millimeter-wave scenarios. Especially when considering computation-sensitive V2X use-cases, for instance, cooperative perception, local HD map acquisition, and cooperative maneuvers, edge computation capabilities are required to locally run some tasks of relevance for the use cases. At the same time, a network connection able to satisfy bitrate, latency, and reliability requirements of communications is needed to guarantee proper connectivity towards the edge cloud. From this point of view, a joint optimization of edge cloud capabilities and mobile network resources might be beneficial. Considering, for instance, a lane merge scenario, vehicles in the area of relevance may be required to frequently send their status information (location, speed, etc.) to the associated access nodes, thus potentially introducing a high volume of data to be handled by the mobile network. The utilization of millimeter-wave radio access technology from this point of view provides beneficial results for offloading the computation tasks close to the access node by using the available close edge cloud capabilities. In 5GCAR, the total computing latency (for tasks to be offloaded) is estimated considering network aspects (transmitting tasks to the edge and receiving the outputs) and computation latency at the edge node. Please note that the above calculation can be extended to consider energy consumption as well. Consequently, 5GCAR considered an optimization model that, according to the amount of tasks scheduled for offloading, is designed with the goal of minimizing the average energy consumption and the overall latency for task completion, where the variables to be optimized are selection of most suitable access nodes considering their current load (this is to limit upload/download time), selection of most suitable edge cloud (this is to limit computation time), access node, and vehicle's transmission power (to limit power consumption).

In order to provide more details about the joint optimization, the assumptions considered in this study are now provided. It is assumed that local access nodes (e.g., BS-type RSUs) are extended with edge computing capabilities. In addition, it is assumed that access nodes and vehicles have sectorized beam pattern antennas. Access nodes can get limited information about road traffic, such as traffic density, from external application servers.

To optimize scheduling and resource allocation of offloading tasks from the vehicles to the associated access nodes, the following procedure is repeated whenever a vehicle triggers the off-loading

process by sending a request to the access node. In this example, the access node is considered as a controller that is responsible for making optimization decisions.

1. A vehicle randomly (according to a distribution) generates computing tasks, some of which are scheduled for offloading. The possible reasons for offloading include vehicle's hardware limitations, lack of real-time traffic and road data at the vehicle, etc.
2. The vehicle's computing tasks' queue is updated with new arrivals and the number of tasks scheduled for offloading.
3. Dynamic channel model between the vehicle and the access node is estimated, taking into account the vehicle's speed and distance from the access node at each time.
4. The total computing latency (for tasks to be offloaded) which consists of the following parts, is estimated.
 - a. Latency of task uploading from the vehicle to the access node (uplink transmission).
 - b. Task execution time at the access node.
 - c. Latency of downloading the computing result from the access node to the vehicle (downlink transmission). (Uplink and downlink transmission latencies depend on the dynamic channel model estimated in the previous step, millimetre bandwidth, transmit powers of the vehicle and the RSU, uplink interference, and noise power)
5. A model is considered for the total communication and computing energy consumption by access node and the vehicle.
6. An optimization model is considered with the objective of minimizing the average energy consumption and some constraints including latency constraint, maximum transmit powers, and stability conditions of the vehicle's computing queue. The variables to be optimized are RSU and the vehicle's transmit power, as well as the number of tasks scheduled for offloading.

The output of the optimization problem determines whether a certain number of tasks can be offloaded to a certain access node.

3.5. Multi Connectivity Cooperation

V2X services are expected to exploit infrastructure-based links (i.e., Uu) as well as direct vehicle-to-vehicle (V2V) links (i.e., sidelink—SL). These two links have different characteristics and consequently are associated with different features. For instance, SL is expected to provide low latency reduction and out-of-coverage support while Uu is expected to offer higher reliability as well as possibilities of connectivity towards remote backends. 5GCAR considered that the joint exploitation of the two communication modes (i.e., Uu and SL) might bring benefits in order to meet the requirements of complex V2X environments. 5GCAR also identified the issue that the selection of a suitable communication mode or technology should not only be driven by the QoS requirements of the related traffic. Indeed, as many V2X use cases (e.g., lane merge) focus more on the completion of a certain action (e.g., a vehicle safely merging into the main road), such type of information should be taken into consideration on the chosen communication mode/technology to allow the completion of the use case, i.e., the management of multi-connectivity options should be carried through the utilization of use case-aware mechanisms.

The benefits deriving from multi-connectivity cooperation are in terms of improved performance (e.g., reliability and data rate), together with higher resilience to link failure (e.g., redundancy).

More details on the analyses conducted by 5GCAR on different options for achieving redundancy of Uu and PC5 links are given in Section 4.6, while an analysis on how to achieve use case-aware multi-link connectivity is given in Section 4.3.

4. Detailed Analysis of Selected Features

This section provides a detailed description of some selected network enhancements introduced by 5GCAR.

4.1. V2X Service Negotiation

The feature of adapting the behaviour of a service according to network capabilities has been studied from different angles, but with the main focus on rate adaptation. The first approach has focused on adapting transmission rate on an end-to-end basis as for instance considered in [18,19], where both the sender and receiver perform some kind of measurement and the receiver transmit feedback to the sender to adapt the end-to-end data rate. Rate adaptation with network support has been also considered by 3GPP in [20], which defines progressive download and dynamic adaptive streaming over HTTP (3GP-DASH) for continuous media delivery. Within 3GP-DASH, server and network-assisted DASH (SAND) introduces messages between DASH clients and DASH-aware network element (DANE) or between various network elements for the purpose to improve the efficiency of streaming sessions by providing information about real-time operational characteristics of networks, servers, proxies, caches as well as DASH client's performance and status. The above-listed approaches, although enabling network-application interaction, only focus on a reduced set of adaptation capabilities (i.e., rate) that might be too limited for V2X use cases. Furthermore, such approaches do not consider the availability of additional information regarding the service at the network as they only focus on the network providing information for rate adaptation. The aspect of interaction between the application and the network has been enhanced by 5GCAR as follows.

The V2X service negotiation represents a network procedure introduced within 5GCAR with the scope to enhance the network awareness about service requirements, which is usually referred to QoS only. Examples of such additional requirements are spatial/time information associated with the service (i.e., the location where the service is expected to run and for how long), information about receiver (i.e., vehicle) status, such as its location, speed, intended trajectory, etc. This set of information enhances network knowledge about the service and can be then exploited by the network to optimize service delivery. In a similar way, the service might also benefit from additional network information, e.g., network capability in fulfilling QoS in a certain area/time or in transferring a message within a certain deadline. As an example, the service can select the most appropriate vehicle driving status (e.g., speed, route, and level of automation) thanks to its increased network awareness. In summary, the benefits introduced by the V2X service negotiation are in terms of: (i) Supporting a more dynamic network-service negotiation with additional information exchanged compared to current QoS-based negotiation procedures; (ii) increasing network awareness on V2X service features which can be used e.g., to optimize the service delivery; (iii) increasing service awareness on network capabilities, e.g., network capabilities on delivering a certain service; and (iv) facilitating the introduction of V2X-specific network features, e.g., the V2X service negotiation gathers and provides service-specific information to other network functions.

An example of an interaction between a V2X service and the V2X service negotiation is depicted in Figure 6. Considering a 5G system, the V2X service can be an AF interacting with a 5G core network, while the V2X Service negotiation can be an enhanced network functionality of e.g., a NEF or of a PCF. From Figure 6, the V2X service provides the network with the V2X service descriptor, which contains the following information:

- Application information, which could be e.g.,:
 - Application type (video, voice, remote control, etc.).
 - QoS, which might the desired QoS by the application (e.g., a certain 3GPP QoS Profile) or a request by the service towards the network receiving information about the QoS the network is expected to support in a certain location.

- Message size, for file-transfer services such as HD map acquisition, software updates, etc.
- Spatial/time constraints, e.g., relevance area (e.g., a certain message should be transferred before the vehicle leaves/enters the relevance area, or the application is happening in a certain location, for instance for a lane merge); message transfer deadline; application duration time (the time interval which the application is expected to last).
- Vehicle information, for each vehicle involved in the application, which could be, for example:
 - Location (e.g., the current location of the vehicle).
 - Speed.
 - Planned trajectory (e.g., waypoints and associated timestamps).

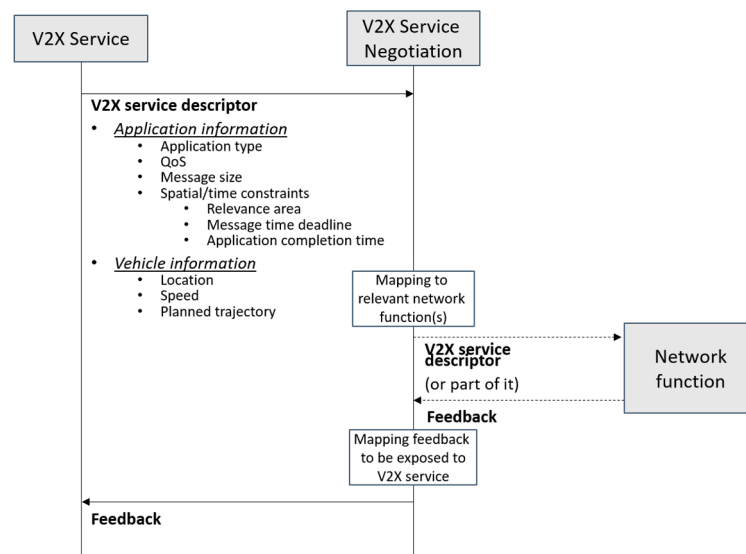


Figure 6. Flow diagram and information exchanged for the vehicle-to-everything (V2X) service negotiation.

According to the V2X service descriptor, the V2X service negotiation maps some relevant network function(s) that receive the descriptor (or part of it). If the V2X service descriptor asked for a request, the V2X service negotiation provides feedback to the service considering the information receiving from the relevant network function(s) contacted upon reception of the service descriptor. The content of the feedback varies depending on the V2X service descriptor, for example:

- A feedback answering a request for information about the QoS the network is expected to support in a certain location.
- A feedback on whether the network expects that the message transfer described in the V2X service descriptor will be completed by the deadline.
- A feedback on whether the network expects to fulfil the desired QoS for the whole application duration time indicated in the descriptor.

An example of V2X service negotiation applied to a file transfer service (e.g., HD map update) is shown in Figure 7. In this example, the V2X service informs the network that a certain amount of data should be transferred towards a certain vehicle within a certain latency budget, also providing information about the expected trajectory of the associated vehicle within the area of relevance of the transfer. The network maps the information provided with the descriptor to internal network information, such as for instance network topology, expected load, and provides a feedback to the service containing information on when to start/pause the delivery in order to support an efficient file delivery within the transfer deadline.

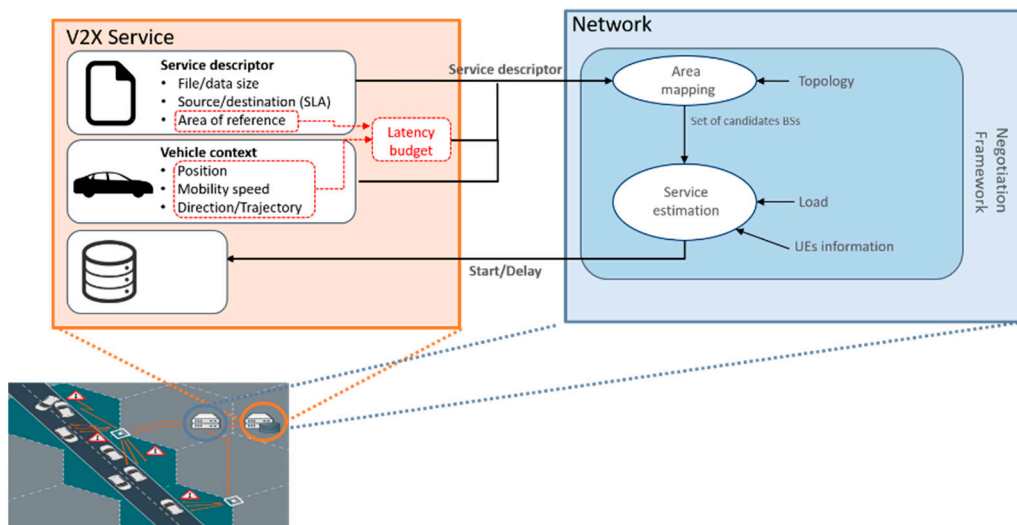


Figure 7. An example of a V2X service negotiation applied to a file transfer service.

4.2. Location-Aware Scheduling

Prior works on scheduling for vehicular environments mainly focused on guaranteeing low-latency and reliable communications, which requires additional efforts compared to legacy environments due to vehicles' mobility. The work in [21] defines a scheduling algorithm running at the RSU based on the definition of a weight taking into consideration channel conditions, vehicle speeds and vehicle and/or data priority. A more detailed list of works on this topic can be found in [22], where the listed works in literature mainly consider that data to be delivered to vehicles are "local". This means that data received by the vehicles will be immediately used once upon reception as they contain information (e.g., positions and speeds of other vehicles) relevant to the area where vehicles are moving in that specific time interval. As considered in 5GCAR, V2X applications deal with geographical data but in some cases (e.g., HD map dissemination) such data might be generated and available before the vehicle approaches the target area. This means that the network should take into account the geographical area of the data when scheduling data delivery, i.e., potentially data transfer can be "relaxed" in this case to avoid congestion in the network. Considering this aspect of "relaxed" or low-priority file transfer, one solution that has been considered relies on adapting the file delivery window to reduce congestion. From this point of view, the most relevant example is represented by TCP_LEDBAT [23] that has been introduced with the goal to provide a "less than best-effort" service delivery for applications that have loose constraints in terms of data delivery deadlines or throughput. Although such approaches represent a valid solution to deliver low-priority content (e.g., software updates) and might potentially enable a reduction in the cost of delivery, they do not take into consideration time or spatial deadline aspects for file delivery. This aspect is of primary importance for some V2X services (e.g., HD map download) where data content should be delivered before the vehicle enters the geographical area the file refers to.

The location-aware scheduling is an enhanced network procedure designed within 5GCAR to efficiently manage the transfer of messages with associated spatial/time deadlines. When mapping a certain service to a certain QoS, the location-aware scheduling considers location information, with regards to both vehicle location information (e.g., expected or planned trajectory) and service location information (e.g., area of relevance of a certain message). This is to take into account the requirements that should be interpreted in terms of geographical region (e.g., the vehicle should receive/transmit the message before entering/leaving a certain area) instead of absolute latency budget. Having this in mind, the transfer of the message could be optimized considering the spatial (or time) deadlines of the message, jointly with the vehicle status information (position, speed, etc.). The location-aware

scheduling then enables a dynamic adaptation of QoS parameters taking into consideration additional V2X service information instead of using static QoS mapping.

In the remainder of this section, we consider a high definition (HD) map acquisition use case, where different messages (layers 1–2 and layer 3) have different areas of relevance, meaning that such messages should be received before entering the associated relevance areas (Figure 8). In addition to the relevance area, additional information can be associated with a message such as the message size. From Figure 8, two different latency budgets for message reception can be extrapolated for delivering layers 1–2 and layers 3, considering the speed of the vehicle and its distance from the relevance areas. These latency budgets might represent an additional source of information for the network to be exploited for optimizing the message delivery. From Figure 8, the location-aware scheduling receives the following inputs from the V2X service:

- Message information: relevance area (or time deadline), message size.
- Vehicle information: location, speed (or planned trajectory with associated timestamps).

The above information is exploited by the location-aware scheduling to provide the following outputs:

- Scheduling information: transfer planning, etc.
- QoS requirements: QoS Profile, etc.

The outputs of location-aware scheduling could be delivered to the service of relevance, e.g., the service might be informed about the transfer planning to be aware of when to trigger the message transfer. On the other hand, some parts of the output could be delivered to the network for different purposes. For instance, information about the transfer planning can be used by network nodes (e.g., RAN or UP nodes) as additional information regarding the expected traffic load. Furthermore, information about the associated QoS requirements can be then used by the network to enforce the correct QoS for the message transfer (e.g., increase priority of transfers for which vehicles are closer to the relevance area).

From Figure 8, considering the information about the vehicle's location and associated speed and trajectory, the location-aware scheduling can schedule in an efficient way the delivery of the layers 1–2 and layer 3. Considering the larger size of layers 1–2 compared to layer 3, together with the information that the relevant area of layers 1–2 is closer to the vehicle with respect to the one of layer 3, the location-aware scheduling schedules the transfer of layers 1–2 and layer 3 in two different windows, in order to guarantee the overall transfer of all layers before the vehicle approaches the associated relevance areas, also guaranteeing that the network load is kept low by splitting and spreading the transfer of the two messages into two different time windows.

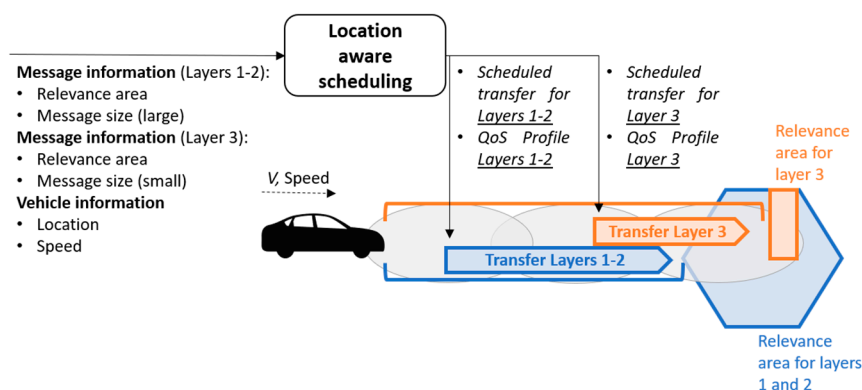


Figure 8. Example of the utilization of location-aware scheduling applied to high definition (HD) map acquisition use case.

4.3. Use Case-Aware Multi-RAT, Multi-Link Connectivity

Given the diversity of use cases, a combination of links and/or RATs that could deliver the requirement of V2X use cases could also be diverse. For example, while cooperative safety is extremely time-sensitive, and the success of cooperative manoeuvre highly depends on persistent coverage, cooperative navigation is in real need of high bandwidth from the network. The selection of the best possible combination of links/RATs is also a changing choice, given the mobility of the vehicles.

Use case-aware multi-RAT, multi-link connectivity is a possible solution designed in 5GCAR to guarantee delivery of use-case specific service requirements (QoS). Due to the high mobility as well as the latency/safety-critical nature of most V2X use-cases, it is very challenging to guarantee QoS delivery for the whole duration of the use-case. Use-case aware, in this context, refers to differences in characteristics and demands of different use-cases. For example, in the class of cooperative safety use-cases, persistent coverage during the time period of the use-case is very important to be guaranteed. While in the class of cooperative perception use cases, providing high data rates might be challenging especially in a highly congested network. The option of having multi-RAT, multi-link connectivity, would provide higher network availability via spanning radio resource pool and exploiting it in two forms multiplexing or diversity gain. Some scenarios are, for instance, simultaneous connection to two BSs (when this type of connection is supported) and sending the same packets (diversity gain) or different packets (multiplexing gain) on both links, or even in case where multi-connectivity is not supported, the possibility of selecting the best link/RAT among all possible options would provide link quality enhancement by exploiting diversity gain.

To be able to use this component, we assume the following: (i) vehicles/UEs are capable of simultaneous connection to more than one link either within one RAT or to different RATs; (ii) communication network support of multi-link/RAT connectivity through information exchange and coordination between protocol stacks of different RATs e.g., LTE, NR; and (iii) a coordinator entity, located within the domain of the MNO the UE is attached to and managing available links/RATs, is responsible for determining the type of multi-link/RAT connectivity for the involved vehicles/UEs.

To evaluate each combination of links/RATs for each use case, in the first step, the concept of “action” should be defined: action is intended as a series of V2X communications with pre-defined QoS requirements to enable successful delivery of the use-case. In this context, “completion of action” refers to the successful delivery of the use-case. Therefore, the evaluation should be based on completion of action rather than per-packet analysis of QoS. For instance, a specific type of connection may fail to deliver QoS requirements for some periods of time during the manoeuvre, but it may still be able to provide enough resources for the manoeuvre completion since small periods of service un-availability would not cause a serious problem.

The second step is to consider the “completion time” of the action. There are two scenarios in this step:

- The use-case requires V2X communication of a pre-defined period of time.
- The completion time of the use-case depends on some parameters such as the current location and speed of vehicles as well as the intended type of multi-link/RAT connectivity. However, there is probably an upper bound limit for the completion time defined by the use-case. If none of the available multi-link/RAT connectivity is able to complete the action before the deadline, the use-case will be declared as unavailable by the coordinator.

In the case of the first scenario, the coordinator does not need to compute the time to completion and thus, can proceed to the third step. However, in the case of the second scenario, the coordinator needs to compute the completion time for each multi-link/RAT connectivity based on automotive conditions as well as the intended method of data splitting between links e.g., exploiting diversity via coding or aggregating bandwidth to increase data rate.

In the third step of the evaluation process, the coordinator decides which links/RATs to attach to for each of the involved vehicles/UEs. Taking into account the following parameters which are

assumed to be known or predicted by the coordinator, the coordinator would be able to predict the resulting QoS of each combination of multi-link/RAT connectivity. Those parameters include:

1. The geographical area of relevance and the related information such as buildings topology (to find the most fitting path-loss model) as well as locations and height of BSs.
2. The expected trajectory of all the vehicles in the area of relevance.
3. The data type (e.g., video, status information, warning messages, etc.) and its related properties such as packet size, arrival distribution, etc.
4. Relative priorities among different use-cases and/or vehicles.
5. Expected congestion and the number of available resources for each RAT based on the current waiting time of queues in bottlenecks and the number of connected UEs.

Then, the coordinator compares the expected QoS for the whole action duration against the required QoS of the use-case and reports the promising candidates for the connection of each vehicle/UE. The final decision between those candidates is made with the aim of congestion control via traffic off-loading to the less-congested networks, while avoiding over-provisioning and waste of resources.

The use case-aware multi-RAT, multi-link connectivity optimization presented by 5GCAR can be further extended considering aspects of the trade-off between reliability and latency, as for instance considered in [15]. In this work, the Euclidean norm theory has been used to propose a reliability and latency joint function, and the relationships between reliability and latency are illustrated via Monte Carlo simulations of 5G autonomous vehicular networks. These aspects can be considered in the selection of the link/RAT to be used for V2X use cases.

4.4. Evolution of Infrastructure-Based Communication for Localised V2X Traffic

In many V2X use cases (e.g., cooperative manoeuvres, sensor information sharing, video sharing, and intra-platoon communication) the data traffic that is exchanged among vehicles (V2V) has localized significance. This means that the communicating vehicles that participate in the same use case are located in the same geographical region and there is no need to access a remote server, (e.g., V2X Application Server, and ITS cloud server), while multiple transmission modes (unicast, broadcast, and multicast) might be required. For localized V2X communications, either the cellular (Uu) interface or the sidelink (PC5) interface could be used considering the radio conditions and the environment where the V2V use case takes place. Specifically, the NR-Uu interface could provide guaranteed QoS (i.e., high reliability and low latency) especially in the case of, e.g., no line-of-sight among communicating vehicles, poor PC5 radio conditions, or high PC5 interference due to vehicles' high density. Nevertheless, existing cellular solutions, based on the Uu interface, may need some updates for supporting in a more efficient way the challenging performance requirements that localized V2X services have, which include the need for fast and guaranteed transmission of localized data.

In the existing 3GPP procedures, the control plane latency for establishing a new bearer is significantly larger compared to the requirements of many V2X use cases [24] since core network entities are involved in the establishment of bearers, which increases the required communication and processing delay. More than 130 ms (control plane latency) are needed for the establishment of required bearers for a group of vehicles that participate in the same V2X service, when the vehicles are not connected to a BS (Radio Resource Control - RRC IDLE state) [25]. If the vehicles are connected to the same BS (i.e., can directly ask for resources) the control plane latency is higher than 80 ms [25]. Thus, these large control plane latency values are not appropriate for many V2X use cases. The introduction of new state models for the radio access networks, called "connected inactive", where both the user equipment and the network maintain context information, enabling the quick and lightweight transition from inactive to active data transmission, could support the fast connection establishment problem [26]. The proposed "connected inactive" could contribute to the reduction of the total control plane latency, but it is not enough since the delay for adding new bearers, especially for services, where a group of devices participates, remains high (i.e., more than 80 ms).

The formation of local end-to-end (E2E) radio data paths over the Uu interface, could be used to enable the fast and guaranteed transmission of localized data traffic among the involved devices, satisfying their QoS requirements and the features of the V2X services, as initially presented in [27]. The “end-to-end” term denotes that the (user plane) radio data paths are established among the involved communicating end devices (i.e., vehicles), while the “local” term denotes that the paths are established via the BSs. Instead of using solutions such as local UPFs providing a local UP end-point instead of reaching the UPF within the core, the focus of this TC is that the nodes of the core network do not participate in the user plane transmissions since the data traffic is localized and handled directly among involved BSs. Local e2e paths via the BS can support different communication modes (unicast, multicast, and broadcast) without the need to interact with other entities such as MBMS. Although, according to the proposed solution the data traffic does not pass through core network nodes, online and offline charging methods could be used and specifically with the support of the charging function (CHF) of the 5G system architecture. For instance, the SMF that could be used for the establishment of the local e2e paths could also trigger the appropriate event and support the collection of charging required charging information [28].

Localized communication through the Uu interface requires the introduction of a data routing/forward function at the BS (e.g., gNB) that transmits the data packets, e.g., among vehicles in a fast and guaranteed way without traversing any core network entity (i.e., user plane). This routing table in the BS maps and connects the uplink (UL) and downlink (DL) radio bearers of different UEs for the formation of the local radio paths and consequently the faster forwarding of localized V2X traffic (user plane latency reduction). According to the type of the traffic, the routing table at the BS undertakes to forward the data packet to one or more UEs in the same or neighboring cells (e.g., multicast and unicast transmissions). In case UEs are attached to different cells, a possible solution that requires further investigations is to use Xn interface for path establishment and for data packet exchange. Figure 9 provides an overview of the involved entities and interfaces.

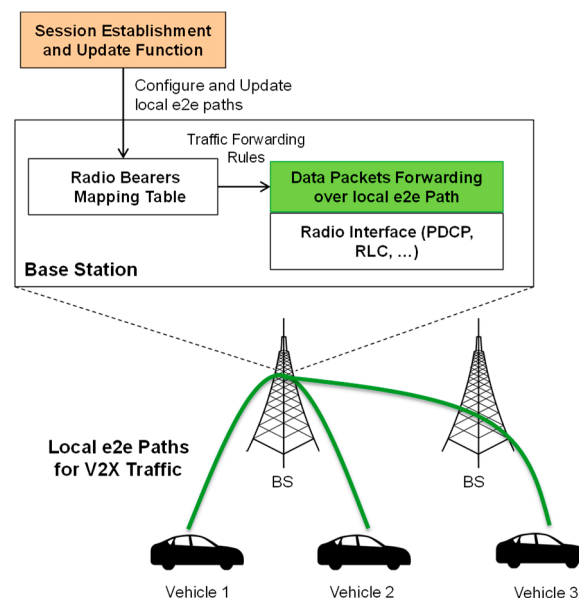


Figure 9. Overview of vehicle-to-vehicle (V2V) paths via the cellular interface.

A UE requests the establishment (or update) of the local cellular V2V paths using RRC, NAS protocols, for localized V2X traffic and to transmit/receive data packets over a local e2e path. The type of the service and the identifiers of other involved UEs in the corresponding V2V service are information that the initiating UE should provide and is used for the establishment of the paths as well as for the configuration of the routing tables. RRC and NAS protocols need an extension to support establishment, update, and release of local cellular V2V paths between the UEs over the gNB(s) as well as to update

and configure the routing table needed for the forwarding of localized data traffic. Based on these RRC or NAS messages, core AMF and SMF can control the establishment, modification, and release of this new type of link (i.e., local cellular V2V paths) as well as to update and configure the routing tables that are introduced at the BSs in order to form V2V paths for localized V2X traffic over the Uu interface.

In general, the low latency and high reliability of multicast and unicast transmissions via the local e2e paths solution allows the realization of V2V services that more demanding and have more complex interactions (e.g., cooperative maneuvers, sensor data exchange, cooperative perception, and see-through).

4.5. Multi-Operator Solutions for V2X Communications

It is a valid assumption that V2X communication is going to be a multi-operator environment since it cannot be guaranteed that only one operator is going to coordinate the vehicular interactions in the overall area. This assumption has been accepted by various organizations [29,30] and is the working assumption in order to make a V2X system operating, under the considered V2X requirements. Multi-operator V2X has implications in the communication via both PC5 and Uu interfaces, as well as in the use of edge computing.

Solutions available in the literature can be based on the use of the already available Uu interface and use of existing data paths, where the data for the two devices attached to different operators need to cross all domains (i.e., access, backhaul, and core) of the two operators, incurring a delay that is not acceptable for most of the cases of V2X communication. On the other hand, the use of SL (i.e., PC5) interface requires coordination among the operators on the use of spectrum resources for ensuring proper reception of the information—without necessarily ensuring high reliability if the communication takes place in an uncoordinated manner (Mode 4 PC5). Other solutions can rely on use of schemes that facilitate the use of both interfaces but require either RAN sharing among the operators or to force the devices to roam from one operator to the other. The first approach which is based on RAN sharing requires agreements in all the coverage areas so as to ensure proper operation of the network whereas the other based on force roaming introduces unacceptable delays for the transition from one operator to the other because of the required attachment procedure. When edge computing is considered the previous problem is even more complex.

Multi-operator V2X communication using PC5 may happen using:

1. Mode 1/3 by applying:
 - Shared carriers, which may face synchronization problems (can be handled with global clock such as GLONASS) and resource overlap. Resource overlap can be solved by applying resource split in TDM way.
 - Separate carriers, which do not face resource collision between different PLMNs but require carrier switching (with interruption and delay) or multiple Rx chains (which increases the cost).
2. Mode 2/4 by applying:
 - Shared carriers, which also may face synchronization aspects which can be handled by applying a global clock such as GLONASS. In this case static configuration is used requiring updates in special cases such as the cross-border situations or when the spectrum pools are reconfigured.
 - Separate carriers, which is based on static configuration—updates may be required in cross border cases. Carrier switching, in this case, requires a well (with interruption and delay) or Multiple Rx Chains (costly).

The above cases should also include the case of PC5 in an unlicensed or licensed spectrum (which would need agreements among MNOs in case of shared carrier). In any case, the configuration should

be static or updates every time a reconfiguration occurs are needed. Also, it has to be stated that the reconfigurations, and the carrier switching come with delays that may hinder meeting the requirements in particular cases. It has to be noted that in case of shared carriers in licensed spectrum agreements between operators are needed.

Multi operator V2X communication using Uu may happen using:

- Typical Uu communication through normal UPF functions (i.e., home routing)
- Local breakout schemes to communicate with vehicles of the other operator. This solution requires implementations with local breakout dependent on the topology (e.g., gNBs in the highway should have access to it) thus increasing the complexity.

When using Uu interface, edge cloud schemes used for processing the data need to provide this information to vehicles of the other operator (e.g., in the case of HD maps use case and vehicles that are associated with different edge clouds owned by an operator). This requires Multi-access Edge Computing (MEC) sharing the same schemes with local breakout solutions described above to reduce the time for sharing this information. Additionally, multicast/groupcast requires duplication of the messages in the two (or more) operators' spectrum.

The previous analysis shows that enhancements are required for dealing with the proper (in terms of delay and reliability) communication of the vehicles belonging to different operators if we want to avoid complex deployments. Additionally, we should consider the case of the cross-country border crossing, where interruptions are going to occur since the UE should register to the other country's operator. The latter procedure requires a significant amount of time (i.e., at least some seconds; it has to be noted that the registration/attach procedure is in the range of 330 ms [31]), and thus a significant interruption in the service, which is not acceptable according to 3GPP requirements, where it is mentioned that "regardless of actual scenarios in place, from passenger experience point of view, automated driving should not be interrupted even when the vehicles cross borders as long as automated driving is permitted by regulation" [32].

The challenges related to increased delays or limited reliability can be solved if the multi-operator problem may be simplified to a single operator based on agreements among operators for a regional split, as for example depicted in Figure 10. Splitting the overall area into regions where only one operator is responsible simplifies the multi-operator environment and enables efficient V2X communication. PC5 communication is efficiently handled by one operator without requiring complex coordination among multiple operators. Also, the edge cloud solutions do not require any further enhancements since a single operator offers such service.

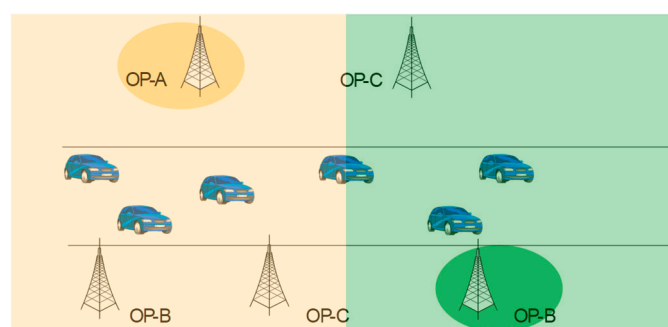


Figure 10. Split of a highway.

In order to minimize the transition time from one operator to another, when crossing the boundaries of an area, the devices will be registered in advance to all available operators. For example, in the case of a vehicle, while performing a typical attachment to its Home Public Land Mobile Network (HPLMN), the subscriber server of the HPLMN triggers through the core network the attachment (i.e., registration of the same vehicle) in all other available operators. Upon attachment to all the networks,

a device is considered to be “connected” to only one of them while being in an “idle” state in the others. To which operator a UE should be connected is dictated by the network; this may be done by direct indication from the operator or via pre-configurations, or the tracking area updates (TAUs). This pre-registration allows to minimize the time needed to steer a device from one operator to another, because it will be RRC connected to the operator who is indicated to serve it and idle for all the other operators in the area under consideration (Figure 11a). It should be noted that the registration to all the operators takes place via interactions among the operators. The selection of through which operator a UE should communicate with the other UEs in the vicinity and the network depends on the geographical location of the device or specific instructions from the network. In the latter case the UE attaches to an operator, who instructs the UE to attach to another UE (Figure 11b). Further enhancements, where the UE may be connected in both operators or connected to the one operator and idle to the other but listening to the downlink signalling channels with higher frequency may further reduce the delays and increase the reliability.

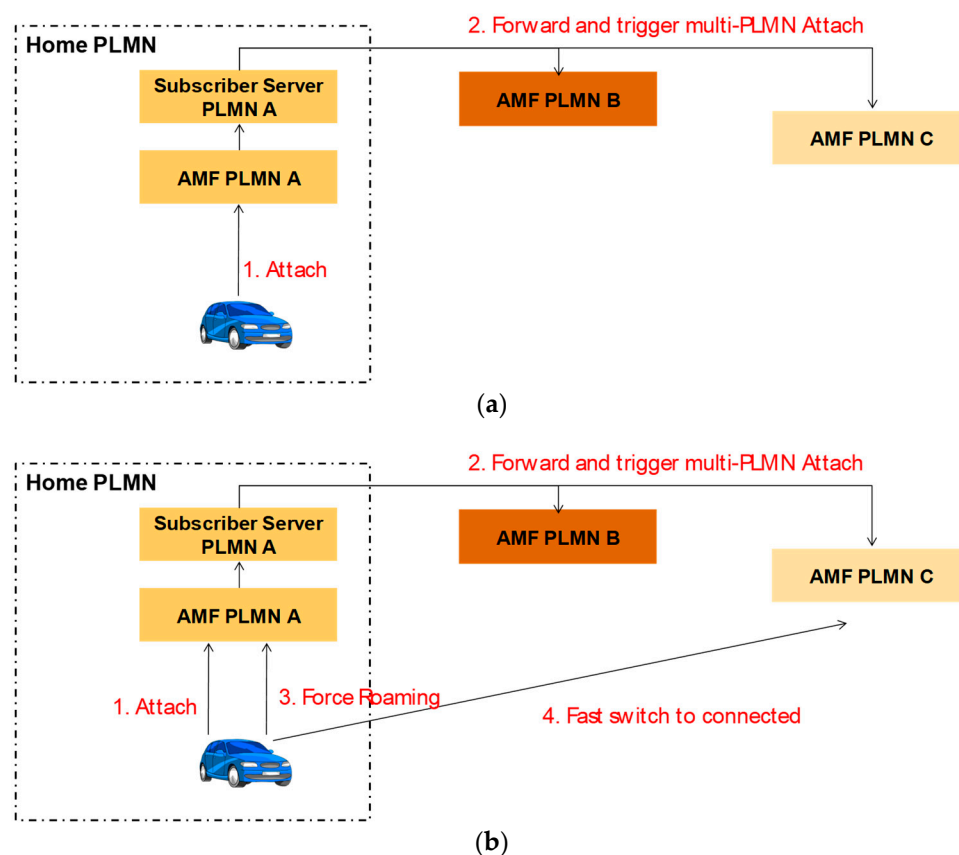


Figure 11. (a) Pre-attach to operators and (b) force roaming procedures.

4.6. Redundant Mode PC5 and Uu

The support of safety-critical automotive applications imposes strict reliability requirements on V2X communications, which for specific use cases needs to be as high as 99.999%. In order to achieve this degree of reliability, both the Uu and PC5 links can simultaneously be used to provide one degree of redundancy [33]. As illustrated in Figure 12, data flows are duplicated and transmitted both on the PC5 and on the Uu links. Since this procedure makes intensive use of resources, its use shall be constrained to specific use cases and scenarios, and limited in space and time, based on specific operator’s policies.

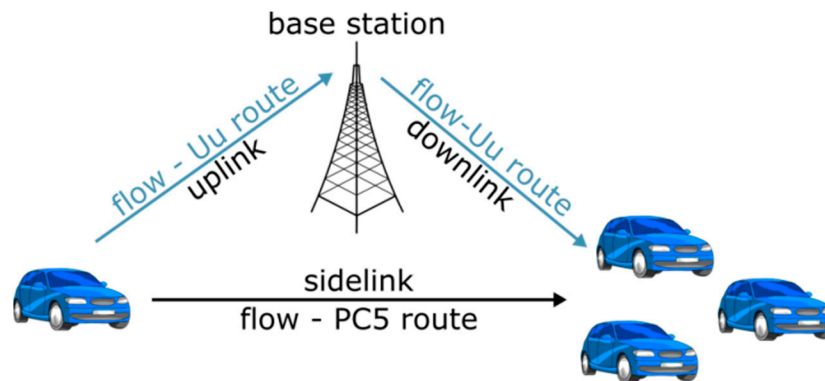


Figure 12. Redundant scheduler, illustrating the Uu and PC5 routes.

In this section, we discuss and compare four different implementation solutions for the redundant scheduler, differentiated by the layer at which the duplication/deduplication of flows is performed. Specifically, the scheduler may operate at the facilities layer, transport layer, RRC layer, and medium access control (MAC) layer.

Among the different solutions proposed, the implementation of the redundant scheduler at the facilities layer is arguably the easiest to implement and to test, and by design it limits the redundancy feature to only a specific subset of applications. Furthermore, the duplication is RAT-agnostic, meaning that flows may be routed over no matter which air interface is available. However, this implementation requires all UEs involved in the communication to implement the ETSI ITS stack [34], and the receiving UEs to be capable of receiving on both interfaces. Being located in a higher layer of the stack, this solution enables no interaction with the lowest layers, not enabling optimizations (such as temporarily stopping the duplication) when one or more links are saturated.

When implemented at the transport layer, the redundant scheduler may rely on already available protocols, such as Multi-Path TCP [35] and quick UDP connections [36]. For this reason, as opposed to the facilities layer implementation, this scheduler may also be made available to other non-vehicular applications running on the same UE, in case of need. On the downside, multipath transport protocols are currently only designed for IP traffic, effectively excluding non-IP flows, which are prominent in V2X. Furthermore, due to its location in the higher layers of the protocol stack, it prevents an optimized use of the duplication when links are saturated, similarly to the facilities layer implementation.

The RRC sub-layer represents another convenient location to implement the redundant scheduler, due to the availability of already standardized components. For instance, RRC is already defined for Release 14 semi-persistent scheduling (SPS), which is defined both for the Uu and the PC5 interfaces. Furthermore, Mode 3 sidelink already enables cross-carrier scheduling, which represents a solid base to define an extension designed to support the Uu + PC5 mode. However, it is difficult to estimate the entity of the modifications to the RRC sub-layer necessary to support the redundant scheduler. Furthermore, it requires the gNB and the RSU to be tightly interconnected, to enable efficient simultaneous usage of both interfaces for a V2I or V2N2I service.

Finally, a redundant scheduler located at the MAC layer may also be based on a concept formalized in Release 14, notably the V2X concurrent inter-band configuration for UU and PC5 communications; the MAC entity implementing the redundant scheduler may be integrated in a similar way to the uplink carrier aggregation. As opposed to higher-layer implementations, the MAC scheduler may control the duplication, and temporally suspend it, in case of channel congestion. On the downside, however, this is the most complex solution to implement: in this case, its location on a lower layer makes it difficult for the scheduler to enable/disable the redundancy for specific flows unless further dedicated cross-layer mechanisms are applied. Furthermore, similarly to the RRC deployment option, it requires the gNB and the RSU to be tightly interconnected, in order to enable efficient simultaneous usage of both interfaces for a V2I or V2N2I service.

5. Conclusions

The 5G system layer architecture will play a fundamental role in the delivery of C-V2X communications capable of providing and sustaining QoS tailored to advanced vehicular applications. In particular, an ensemble of notable feature is highlighted in 5GCAR as critical for V2X communications. In this paper, we provided an overview of the 5GCAR architecture design, along with deeper insights on selected components, which we argue are important for the deployment of V2X.

V2X service negotiation and location-aware scheduling provide the network with insights about the mobility pattern and application requirements of the vehicular UEs, enabling the network to optimize the delivery of services in space and time, with positive repercussion on the resource utilization efficiency, and therefore of the availability and scalability of automotive applications.

Evolution of infrastructure-based communication for localized V2X traffic deals with the exigence of safety-critical traffic to be treated in the shortest possible time, reducing to a minimum the communication latency by introducing routing paths through the infrastructure that run as close as possible to the vehicular UEs.

Multi-link multi-RAT joint exploitation is a key feature to support and facilitate vehicular communications in multiple different ways, providing enhanced reliability jointly exploiting the PC5 and Uu links, increased data rates, and enabling the network to selectively offload different traffic flows on the appropriate links based on the application needs and on the network load.

Multi-operator support is necessary to enable seamless and low latency communication between vehicles and road users in general, regardless of the MNO each single actor is subscribed to. In 5GCAR, approaches and related procedures have been proposed to address the issue.

The integration of the architectural modules developed in 5GCAR, and notably of those presented in this paper are the key for 5G networks to deliver V2X communications capable of satisfying the stringent requirements of safety-critical applications and to provide extra value to the whole automotive vertical by means of advanced cooperative traffic optimization.

Author Contributions: Conceptualization, M.C., L.G., L.M., A.K., P.S., M.M., T.M.; formal analysis, M.C., L.G., L.M., A.K., P.S., M.M., T.M.; investigation, M.C., L.G., L.M., A.K., P.S., M.M., T.M.; writing M.C., L.G., L.M., A.K., P.S., M.M., T.M.

Funding: This work has been performed in the framework of the H2020 project 5GCAR co-funded by the EU. The views expressed are those of the authors and do not necessarily represent the project. The consortium is not liable for any use that may be made of any of the information contained therein.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fallgren, M.; Dillinger, M.; Alonso-Zarate, J.; Boban, M.; Abbas, T.; Manolakis, K.; Mahmoodi, T.; Svensson, T.; Laya, A.; Vilalta, R. Fifth-Generation Technologies for the Connected Car: Capable Systems for Vehicle-to-Anything Communications. *IEEE Veh. Technol. Mag.* **2018**, *13*, 28–38. [CrossRef]
2. Pätzold, M. Fifth-Generation Technology Offers Trillion-Dollar Business Opportunities [Mobile Radio]. *IEEE Veh. Technol. Mag.* **2017**, *12*, 4–11. [CrossRef]
3. 5GCAR. 5G Communication Automotive Research and Innovation. Available online: <https://5gcar.eu/> (accessed on 5 July 2019).
4. 5GCAR. Deliverable D2.1, Scenarios, Use Cases, Requirements and KPIs. Available online: http://5gcar.eu/wp-content/uploads/2017/05/5GCAR_D2.1_v1.0.pdf (accessed on 5 July 2019).
5. 5GCAR. Deliverable D5.1, Demonstration Guidelines. Available online: http://5gcar.eu/wp-content/uploads/2018/08/5GCAR_D5.1_v1.0.pdf (accessed on 5 July 2019).
6. 3GPP TS 23.501. Technical Specification Group Services and System Aspects. System Architecture for the 5G System. Stage 2 (Release 15). Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3144> (accessed on 5 July 2019).

7. 3GPP TS 23.502. Technical Specification Group Services and System Aspects. Procedures for the 5G System. Stage 2 (Release 15). Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3145> (accessed on 5 July 2019).
8. 5GCAR. Deliverable D4.1, Initial Design of 5G V2X System Level Architecture and Security Framework. Available online: https://5gcar.eu/wp-content/uploads/2018/08/5GCAR_D4.1_v1.0.pdf (accessed on 5 July 2019).
9. 3GPP TS 36.300. Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN). Overall Description. Stage 2. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2430> (accessed on 5 July 2019).
10. Ashtikar, S.; Barker, C.J.; Casper, D.; Clem, B.; Fichadia, P.; Krupin, V.; Louie, K.; Malhotra, G.; Nielsen, D.; Simpson, N.; et al. OPEN DATA CENTER ALLIANCE Best Practices: Architecting Cloud-Aware Applications Rev. 1.0. Available online: <http://www.oaca-project.org/wp-content/uploads/2018/07/Architecting-Cloud-Aware-Applications-Best-Practices-Rev-1.0.pdf> (accessed on 5 July 2019).
11. OpenStack. Tricircle. Available online: <https://wiki.openstack.org/wiki/Tricircle> (accessed on 8 July 2019).
12. Slim, F.; Guillemin, F.; Gravey, A.; Hadjadj-Aoul, Y. Towards a dynamic adaptive placement of virtual network functions under ONAP. In Proceedings of the IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Berlin, Germany, 6–8 November 2017.
13. Slim, F.; Guillemin, F.; Hadjadj-Aoul, Y. CLOSE: A costless service offloading strategy for distributed edge cloud. In Proceedings of the 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 12–15 January 2018.
14. ETSI. GS NFV-INF 004; V1.1.1. Network Functions Virtualisation (NFV); Infrastructure; Hypervisor Domain; ETSI: Sophia Antipolis, France, 2015.
15. Ge, X. Ultra-Reliable Low-Latency Communications in Autonomous Vehicular Networks. *IEEE Trans. Veh. Technol.* **2019**, *68*, 5005–5016. [CrossRef]
16. Ge, X.; Ye, J.; Yang, Y.; Li, Q. User Mobility Evaluation for 5G Small Cell Networks Based on Individual Mobility Model. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 528–541. [CrossRef]
17. Feng, J.; Liu, Z.; Wu, C.; Ji, Y. Mobile Edge Computing for the Internet of Vehicles: Offloading Framework and Job Scheduling. *IEEE Veh. Technol. Mag.* **2019**, *14*, 28–36. [CrossRef]
18. Alvestrand, H.; Lundin, H.; Holmer, S. A Google Congestion Control Algorithm for Real-Time Communication on the World Wide Web. Tech. Rep. 3. IETF. 2012. Available online: <https://tools.ietf.org/id/draft-alvestrand-rtcweb-congestion-02.html> (accessed on 5 July 2019).
19. Johansson, I. Self-clocked rate adaptation for conversational video in LTE. In Proceedings of the ACM SIGCOMM Capacity Sharing Workshop, Chicago, IL, USA, 18 August 2014.
20. 3GPP. Dynamic Adaptive Streaming over HTTP (3GP-DASH) (Release 15). TS 26.274. December 2017. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1444> (accessed on 5 July 2019).
21. Zhang, R.; Cheng, X.; Yang, L.; Shen, X.; Jiao, B. A Novel Centralized TDMA-Based Scheduling Protocol for Vehicular Networks. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 411–416. [CrossRef]
22. Sami, M.; Noordin, N.K.; Khabazian, M.; Hashim, F.; Subramaniam, S. A Survey and Taxonomy on Medium Access Control Strategies for Cooperative Communication in Wireless Networks: Research Issues and Challenges. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 2493–2521. [CrossRef]
23. Ros, D.; Welzl, M. Assessing LEDBAT's Delay Impact. *IEEE Commun. Lett.* **2013**, *17*, 1044–1047. [CrossRef]
24. 3GPP Technical Specification. TS 22.186 Technical Specification Group Services and System Aspects; Enhancement of 3GPP Support for V2X Scenarios; Stage 1 (Release 15) v15.2.0. September 2017. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3180> (accessed on 5 July 2019).
25. 3GPP TS 36.331. V14.2.1. Evolved Universal Terrestrial Radio Access (E-UTRA). Radio Resource Control (RRC); Protocol Specification (Release 14). September 2017. Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2440> (accessed on 5 July 2019).
26. Da Silva, I.L.; Mildh, G.; Säily, M.; Hailu, S. A novel state model for 5G Radio Access Networks. In Proceedings of the IEEE International Conference on Communications Workshops (ICC), Kuala Lumpur, Malaysia, 23–27 May 2016; pp. 632–637.

27. Kousaridas, A.; Zhou, C. Local End-to-End Paths for Low Latency Vehicular Communication. In Proceedings of the IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, Portugal, 3–6 June 2018.
28. 3GPP. TS 32.255 Telecommunication Management, Charging Management. 5G Data Connectivity Domain Charging (Release 15). Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3410> (accessed on 5 July 2019).
29. 5GAA. White Paper on C-V2X Conclusions based on Evaluation of Available Architectural Options. Available online: <http://5gaa.org/news/5gaa-releases-white-paper-on-c-v2x-conclusions-based-on-evaluation-of-available-architectural-options/> (accessed on 8 July 2019).
30. EATA. European Automotive and Telecom Alliance. In Proceedings of the 12th European Congress, ITS Beyond Borders, Strasbourg, France, 19–22 June 2017.
31. Trivisonno, R.; Guerzoni, R.; Vaishnavi, I.; Soldani, D. Towards Zero Latency Software Defined 5G Networks. In Proceedings of the IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015.
32. 3GPP. TR 22.886 Study on Enhancement of 3GPP Support for 5G V2X Services (Release 16). Available online: <https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3108> (accessed on 5 July 2019).
33. Lianghai, J.; Weinand, A.; Han, B.; Schotten, H. Multi-RATs support to improve V2X communication. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 15–18 April 2018.
34. ETSI. *EN 302 665 Intelligent Transport Systems (ITS), Communications Architecture v1.1.1*. 2010; ETSI: Sophia Antipolis, France, 2010.
35. IETF. *RFC 6824 TCP Extensions for Multipath Operation with Multiple Addresses*; IETF: Sophia Antipolis, France, 2013.
36. IETF. *QUIC: A UDP-Based Multiplexed and Secure Transport*; IETF: Sophia Antipolis, France, 2019.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).