On Alleviating Cell Overload in Vehicular Scenarios

Martín Trullenque Ortiz[1], Oriol Sallent[2], Daniel Camps-Mur[1], Josep Escrig[1], Carlos Herranz-Claveras[1], Jad Nasreddine[1] and Jordi Pérez-Romero[2]

[1)] i2CAT Foundation, 08034, Barcelona, Spain

[2)] Signal Theory and Communications Dept., Universitat Politècnica de Catalunya (UPC), 08034, Barcelona, Spain

[martin.trullenque, daniel.camps, josep.escrig, carlos.herranz, jad.nasreddine]@i2cat.net, sallent@tsc.upc.edu, jordi.perez-romero@upc.edu

Abstract—**Fifth Generation (5G) networks will support countless new applications and new business models. One of the 5G paradigms is network slicing, which enables the integration of multiple logical networks each one tailored to the requirements of the different services that can be provided by both network operators and vertical industries. One of the services where 5G is expected to have a greatest impact is vehicular-to-everything (V2X) communications, which will have their stringent latency requirements now met. However, the mobility associated to vehicles can lead to cell overload compromising the required quality of service (QoS). To address this problem, in this paper we propose and evaluate the performance of three network overload alleviation techniques to control network congestion provoked by traffic jams using realistic vehicular traces in a network slicing environment. Firstly, we describe the architecture supporting V2X communications. Secondly, the network congestion control approaches are explained. Finally, after providing a complete description of the considered scenario, results will be detailed, showing that the network overload appearing during rush hour can be significantly reduced.**

Keywords— V2X communications, network overload, realistic scenarios.

I. INTRODUCTION

The most significant change from previous generation of wireless communications to 5G networks has been replacing the "one size fits all" paradigm to an ecosystem that simultaneously supports a wide variety of different requirements, from unprecedented high data rates to millisecond latencies. This will be a key enabler for a wide range of applications scenarios and business models, which will enable vertical industries to move from dedicated solutions to a more cost efficient, interoperable, and open ecosystem enabled solution platform.

One of the key enablers to allow the required versatility is network slicing, which allows 5G system operators to compose and manage dedicated logical networks with specific functionality on top of a common shared physical infrastructure. The greater elasticity brought about by network slicing will help to address the cost, efficiency, and flexibility requirements imposed by the large variety of industrial vertical services. Moreover, network slicing will help new services and new requirements to be quickly addressed, according to the needs of the industries, i.e., a faster Time to Market [1].

One of the wide area services that public 5G networks are called to manage is vehicular-to-everything (V2X) communications. In this domain, telecommunication and automotive industries have joined together forming the 5G Automotive Association (5GAA) [2], an industry association that aims to address connectivity needs arising from cooperative connected and automated mobility (CCAM). To help mobile network operators (MNO) to provide V2X services, the 3rd Generation Partnership Project (3GPP) has standardized a V2X slice type in [3] together with the enablers to support it.

Due to the inherent mobility associated to V2X services, managing vehicular networks is particularly challenging. In this respect, many contributions have been done in different fields. Zhang et al. proposed in [4] a deep reinforcement learning approach to perform resource allocation in cellular V2X communications, where high bandwidth demanding traffic is served using vehicle-to-Infrastructure (V2I) while safety related messages are sent through vehicle-to-vehicle (V2V) communications. Luoto et al. created in [5] vehicle clusters in cell edges to prevent running out of radio resources when vehicle density increases. Similarly, in [6] Khan et al. partition a vehicular network in three different slices depending on each vehicle SNR to ensure seamless QoS for video streaming for all of them.

The difficulty to deal with V2X traffic increases when network slicing is considered, as network capacity needs to be shared among slices. This has been an exploited research topic. The authors of [4] and [7] focused on simultaneously serving different vehicular applications. However, none of the previous papers studied the impact that high road occupancy had on the overall performance. Moreover, network capacity sharing between a vehicular service and enhanced mobile broadband (eMBB) users in a single cell scenario was studied in [8].

One of the main roadblocks for research in V2X communications is the modelling of vehicles' mobility. In [4] and [8] vehicles are assumed to enter in the cell or network following a Poisson process. In turn, in [5] and [6] this is tackled through highway synthetic models. Instead, considering realistic datasets enables gaining deeper insight into V2X scenarios and identifying practical situations that may pose challenges on the way that the scarce radio resources are managed. In this respect, the works in [9] and [10] provide public and realistic datasets of the cities of Cologne and Luxembourg respectively. Using the Cologne dataset Uppoor and Fiore performed a macroscopic and microscopic analysis of vehicular mobility in [11]. However, the analysis focused on metrics such as the dwell time and traffic flow across base stations over time and did not consider the dwell time between neighbouring BS pairs, which is of relevance when detecting events such as traffic jams. Moreover, the implications the traffic had in terms of radio resources were not discussed.

To the best of our knowledge, our recent paper [12] was the first attempt to address the problem of V2X traffic overload in a realistic 5G deployment scenario, studying a traffic jam situation occurring in Cologne's scenario. However, exploiting the knowledge extracted from the vehicles 'mobility analysis to perform an efficient load balancing approach was left as future work.

Leveraging the detailed analysis of vehicles' mobility conducted in [12], this paper presents a step forward towards a systematic characterisation of cell overload in vehicular scenarios. Beyond the formulation of case-specific solutions, the ultimate objective is to devise a comprehensive perspective on cell overload alleviation strategies of general applicability. First, we assess the impact that network congestion control mechanisms applied at the vehicular application layer have on the overall network load. In this respect, the paper focuses on modifying the frequency at which vehicular packets are sent when the network is severely congested. Second, we propose and evaluate a load balancing strategy to alleviate a network overload. The main novelty in the proposed solution is that it exploits the knowledge acquired from the analysis of the traffic jam situation to come up with a more efficient load balancing. Finally, we bring together the benefits from the two contributions to assess the network performance if sharing the capacity between different slices.

The remainder of the paper is as follows. Firstly, Section II describes the approach that 3GPP has taken to accommodate V2X services. Subsequently, Section III focuses on describing different overload alleviation approaches. Section IV will describe the considered scenario and present the results obtained by applying the cell overload alleviation strategies. Finally, we draw the paper conclusions as well as point to future work to be done in Section V.

II. V2X SERVICES IN 5G NETWORKS

To effectively support V2X services over mobile networks, 3GPP has taken a dual approach. First, it has defined a set of network enablers that facilitate the deployment of V2X services. Then, it has defined a set of application enablers to ease the integration of V2X application functions with the 5G network.

Regarding V2X network enablers, 3GPP TS 23.285 [13] defines a set of architectural enhancements, including a PC5 interface for direct V2V communications, an Uu interface for communication between vehicle and the network (V2N), inter-PLMN (Public Land Mobile Network) integration mechanisms to deliver V2X services spanning multiple public networks, QoS enhancements for V2X, and a multimedia broadcast/multicast service (MBMS).

Regarding V2X application enablers, 3GPP TS 23.286 [14] defines a V2X application enabler (VAE) function that abstracts all the previous complex network capabilities into an API that is easy to consume by the V2X application function (AF). The V2X AF is commonly implemented as a V2X gateway component that includes an ETSI G5 facilities layer [15], as well as the ETSI G5 V2X applications.

Among the different ETSI G5 services available cooperative awareness message (CAM) makes sure that all the road users and roadside infrastructure are informed about each other's position, dynamics and attributes [16]. In this respect, CAMs are a key enabler of many applications, such as cooperative adaptive cruise control [17] or safety applications, as described in subsection 5.4 of ETSI TR 102 698 [18]. The construction, management and processing of CAMs is done in the facilities layer within the ITS communication architecture defined in ETSI EN 302 665 [19]. According to the ITS-G5A network defined in ETSI EN 302 663 [20] CAM messages originated by a certain vehicle are received by all the vehicles within a certain downlink range. Furthermore, among the multiple management functions, the transmission interval limits between CAM messages are within 100ms and 1s. This interval time is configurable and adaptive. It must be noted that the greater the message rates are, the more accurate measurements will be reported at the risk of potentially overloading the radio channel.

Following the 3GPP proposed architecture, Fig. 1 depicts an exemplary V2X service architecture based on the Uu interface, a VAE server and a centralized V2X AF that processes all CAM messages from the vehicles in the network. The V2X AF redistributes the received CAMs in the uplink across a specific geographical area, which is considered a parameter of the V2X service. The V2X AF uses the VAE server to resolve the IP address of the UEs, which can be attached to different base stations. Moreover, the VAE also indicates to each on board unit of a vehicle the message rate in order to avoid overloading the channel. In our analysis we consider that the mobile network is not MBMS capable since this feature is not widely supported in current public networks. Furthermore, only V2X Uu interface is considered.

III. CELL OVERLOAD ALLEVIATION STRATEGIES

From the perspective of the cellular network, traffic jams are particularly demanding in terms of the generated V2X traffic load, since the cell's load will be affected not only by the increase in the number of vehicles or by their larger dwell time but also by the V2X applications' characteristics. Indeed, given that a CAM message originated in a certain vehicle is forwarded in downlink to all vehicles within a certain area, the number of downlink CAM messages dramatically increases during a traffic jam. Although this effect would be alleviated with the use of MBMS, this feature is not widely available in current systems, so this paper will intend to propose a solution for supporting V2X services in the absence of MBMS.

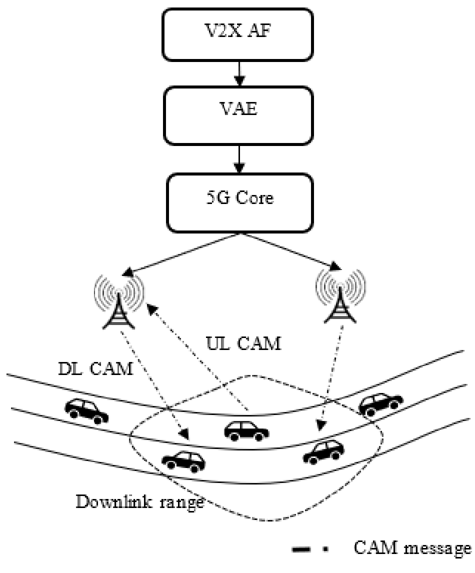

Fig. 1: Architecture of modelled service.

Thus, considering that cell overload in vehicular scenarios is likely to appear, this section is devoted to discuss different mechanisms that can contribute to alleviate it. Firstly, we exploit the capabilities that V2X application layer brings to adapt the CAM message rate to the current load. Secondly, a mobility load balancing approach exploiting vehicle's mobility is considered. Finally, the capabilities of capacity sharing among radio access network (RAN) slices will also be leveraged.

A. Alleviation through V2X application layer

The multiplicative effect traffic jams have in terms of CAM messages arises the need of adjusting the CAM message rate to reduce the usage of the downlink radio resources. The decentralized congestion control (DCC) algorithm, defined by ETSI in TS 102 687 [21], has been widely exploited within the framework of LTE V2V communications in [22] using synthetic data or in [23] using the TAPAS Cologne dataset. However, none of these studies focused on facing significant variations of road occupancy.

The DCC algorithm is jointly controlled at MAC, network and facilities layer. It adjusts the frequency of the CAM messages depending on how congested the radio link is. The channel congestion is determined monitoring the fraction of occupied resource blocks (RB) in a cell. Then, using this congestion metric, the CAM message rate is adjusted between 100ms and 1 second as defined in the annex C.2 of ETSI TS 103 175 [24].

Rather than having a dynamic CAM message rate, we assess the impact that different CAM message rates have both in terms of demanded load and quality of experience (QoE), since it allows a better understanding of the results. While high message rates provide constant updates of surrounding vehicles position at the price of potentially overloading the network, large time between CAM packets can introduce position errors when estimating neighbours position, which might negatively impact certain applications (e.g., cooperative adaptive cruise control). To this end, we define the average estimate position error when the cell is not overloaded as

$$e = \sum_{k=1}^{K} v_k CAM\_interval, \qquad (1)$$

where K denotes the number of connected vehicles to a cell, v_k the k^{th} vehicle speed and $CAM\_interval$ the interarrival CAM packet time. In the event of having an overload, the average estimated position is defined as

$$e = \sum_{k=1}^{K} v_k CAM\_interval * \rho, \qquad (2)$$

where ρ is the demanded load, defined as the ratio between the total number of required RBs and the number of available RBs.

B. Alleviation through Mobility Load Balancing

Traffic jams typically present an exacerbated increase of vehicles density in a particular area, which can be assessed by performing an accurate vehicular mobility analysis [12]. In this respect, if the traffic jam is located close to a cell edge and users exhibit low reference signal received power (RSRP) to their connected cell, handover offsets can be adjusted to divert these users to their neighbouring cells.

This has been a common approach to relieve heavily congested cells and has been evaluated using live data [25] or within the scope of vehicular communications with synthetic vehicular traces [26]. Furthermore, this might be automated by exploiting self-optimizing network capabilities such as MLB (mobility load balancing) [27], which aims to optimize cell handover parameters to cope with the unequal traffic load.

By implementing MLB, a connected user to cell BS_i will now be forced to reconnect to BS_j if it satisfies the following inequality

$$RSRP_{BSj} > RSRP_{BSi} - \beta \qquad (3)$$

, where $RSRP_{BSi}$ and $RSRP_{BSj}$ denote respectively the RSRP from BS_i and BS_j, and β is the handover offset in dB that we introduce. Similarly, users connected to BS_j will not connect to BS_i until the following inequality is fulfilled:

$$RSRP_{BSj} < RSRP_{BSi} - \beta\,. \qquad (4)$$

C. Alleviation through dynamic capacity sharing in RAN slicing

The physical resource allocation across different slices is a vital step in network slicing, since failing at performing a correct split of the available resources may lead to situations such as oversizing slices that do not demand a lot of resources while not providing enough resources to others, failing its services to operate correctly.

Moreover, in the context of vehicular communications the way these resources are shared is critical, since the inherent mobility associated with vehicles may pose different situations on vehicular resources demand. While traffic jams may lead to a high demand of network resources, in low road occupancy conditions services in other slices may benefit from the lower demand of resources coming from vehicular services.

Furthermore, in the event of not having enough resources to meet all the slices 'requirements (i.e., during network overload) the resource allocation to each slice will determine which service observes their requirements degraded. Within this framework, it is crucial to evaluate the importance of each service. While for critical services as vehicular communications failing to meet their requirements may lead to the service interruption, other services such as generic Internet users may continue to operate even under some performance degradation (e.g. experiencing some video quality reduction or some additional delay in internet access).

Gathering these two points, in this paper we analyse the impact of two resource allocation strategies. First, we assign to each slice the average resources they require. Then, since vehicular communications is a critical service, we assign a higher number of resources to this slice to absorb the load increase caused by a traffic jam.

IV. PERFORMANCE EVALUATION

This section firstly describes the considered scenario followed by a description of the RAN infrastructure and its configuration, the deployed slices and corresponding services. Subsequently, the evolution of the demanded load will be assessed. Finally, the impact of the various overload alleviation approaches is depicted.

A. Scenario description

The evaluation scenario consists on a 12 sq. km urban area placed in the city of Cologne (Germany). The considered radio access network is composed by a set of 16 cell sites as depicted in Fig. 2. The location of the cell sites is as defined in the Telekom network available in [28]. The area of study is the central part of Fig. 2. [29], while cell sites at the edge (e.g., BS12, BS14, BS15) are taken into consideration to limit the border effects.

The vehicular traces have been obtained from the Cologne city lab project, available in [9]. Out of the whole vehicular dataset of Cologne, this work studies the recording taken in a 7h time frame (from 13:05 until 20:05). Within this time frame, road occupancy experiences significant variations, ranging from fluid traffic to high road occupancy. There are over 40.000 different vehicles driving in the interest area, with a maximum of 450 simultaneously in the highest road occupancy moment, which takes place approximately at 18:00.

B. RAN deployment

We consider that each cell site or Base Station (BS) operates a single cell centred at 2.1GHz band, using frequency-division duplexing, with 20MHz channel bandwidth for each uplink and downlink. To this end, the selected a subcarrier spacing (SCS) is 15kHz, according to Section 5.3 of [30]. With these values of SCS and channel bandwidth, the transmission bandwidth includes 106 RBs, each one consisting of 12 consecutive subcarriers in the frequency domain. For normal cyclic prefix, 14 OFDM symbols can be transmitted per slot and per subcarrier.

The path loss model chosen is the urban area model as defined by 3GPP in Section 4.5.2 of [31]. No slow fading term has been considered. The transmitted power at each base station (BS) has been adjusted to have a data rate around 800 kbps at the cell edge. The rest of the parameters have been chosen from Annex C of [31] and are summarised in Table I.

Each user connects to the cell with lowest path loss. For a user experiencing a given signal to noise ratio (SNR), we consider Shannon's bound to derive the achievable spectral efficiency in the radio link (i.e., $\log_2(1+SNR)$ bits/s/Hz). Then, we consider that the modulation order (Q) and code rate (R) used in the radio link are properly selected for these radio conditions. To this end, Section 5.1.3.1 in [32] is considered to map the achievable spectral efficiency to a MCS index. As a result, the number of bits that a given user is able to transmit on 1 RB during 1 slot is $12\times14\times Q\times R$ bits.

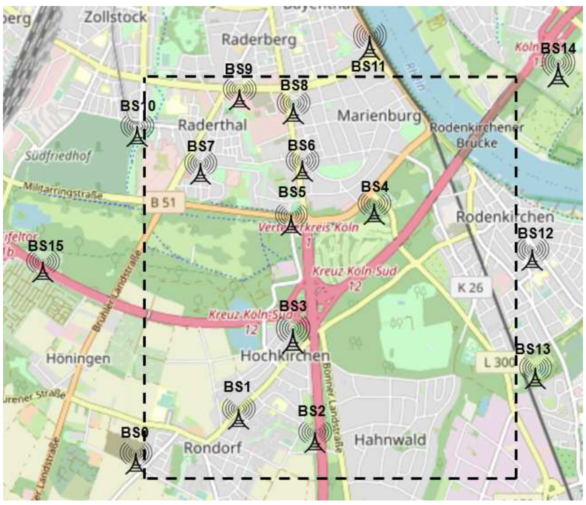

Fig. 2: Considered scenario in Cologne.

TABLE I: RADIO PARAMETERS.

BS parameter	Value
Antenna height	30m
Noise figure	9 dB
Antenna gain	15dB
UE parameters	**Value**
Maximum transmitted power	21dBm
Antenna gain	9 dB
Noise figure	9 dB

TABLE II: CONSIDERED TRAFFIC MODELS

V2X slice	Value
Packet size	300B with probability 1/5 170B with probability 4/5
Downlink message area	1.6 10^5 sq. m
mMTC slice	**Value**
Number of users	150
Guaranteed data rate	3Mbps
eMBB slice	**Value**
Number of users	2500
Uplink target data rate	1Mbps
Uplink sessions percentage	20%
Downlink target data rate	2.5Mbps
Downlink sessions percentage	80%
Av session duration	15 seconds
Session generation rate	12sess/user/h

C. Slices and services

The considered services attempt to replicate the automotive stakeholders defined in Section 6 of [33]. With this objective, three slices have been defined. eMBB serves generic Internet traffic. Massive machine-type communications (mMTC) replicates traffic coming from cameras belonging to a road traffic authority while V2X mimics a service that can be provided by an MNO to a vehicle manufacturer.

The eMBB slice carries generic Internet traffic. 2500 background users uniformly distributed within the area and moving at a constant speed of 3 Km/h as in [34] are considered. These users generate traffic on a session basis as defined in Table II.

The mMTC slice only carries uplink traffic generated from surveillance cameras. These cameras generate a constant data flow and require a guaranteed data rate. Their position has been spread uniformly in the targeted scenario and they are assumed to generate High-Definition video quality constantly.

Finally, the considered V2X slice delivers CAM messages as described in Section II. Different interarrival times are considered in order to define the associated network load impact throughout the considered time frame. The packet size modelling is taken from [35]. The complete description is defined in Table II.

D. Offered load

Having detailed the scenario, this subsection intends to assess the generated load in the described scenario.

As per our analysis in [12], the focus is placed on BS3 of Fig. 2, as it has the largest fluctuation of road occupancy ranging from 22 connected vehicles under low road occupancy conditions to more than 210 when roads are congested, at approximately 18:00. This is due to its position in the target area, supporting traffic from two highways and being close to a junction, which also exposes the traffic variations of vehicles exiting a highway and merging to another. Particularly, traffic to BS3 mainly comes from BS4, BS5 and BS15 and this traffic increase is coming from either the cross highway (cars entering from BS4 and BS15) or the cars entering from the Cologne city centre (North).

To gain further insights regarding the generated loads distribution, Fig. 3 illustrates in blue the total demanded load (from all slices) in BS3 over time. This is defined as the ratio between the total number of required RBs and the number of available RBs. The figure also shows in red the load contribution associated to the V2X slice. In turn, Fig. 4 presents the total demanded load for BS3's neighbours. A static CAM interarrival interval of 500ms as in [12] has been considered. The absence of network overloads motivates the implementation of a load balancing strategy to force BS3 users placed at the cell edge to reconnect to neighbouring BSs.

The increase in vehicular traffic results in an overload situation in BS3 starting at approximately 16:00. Applying the same methodology as described in [12] we observe that the traffic jam is mostly associated to vehicles following the path BS5→BS3→BS4, which experience an average dwell time increase in BS3 from 58 s at 2PM to 950 s at 6PM. Moreover, given that a CAM message originated in a certain vehicle is forwarded in downlink to all vehicles within a certain area, a multiplicative impact on the number of downlink CAM messages appears during a traffic jam.

E. Performance of the cell overload alleviation strategies

Having assessed the impact of vehicles' mobility and associated services in terms of generated traffic load, this subsection intends to illustrate the performance of the load alleviation strategies detailed in Section III. Firstly, a stand-alone evaluation of each strategy is performed in order to assess at what extent each mechanism can contribute to alleviate the overload. Then, insights on how the various mechanisms could operate synergistically are provided.

Regarding overload alleviation at application layer, we evaluate in Fig. 5 the impact on the total load when applying different CAM message intervals (200ms, 500ms, and 1s) on BS3. It can be observed that since under low road occupancy conditions the load associated to the vehicular slice is irrelevant, higher message rates can be safely used. However, when roads get congested, working with CAM interarrival times lower than 1 second leads to a significant overload.

Beyond the cell overload alleviation performance, in Fig. 6 we assess the impact different CAM message rates have in terms of position estimate using (1) and (2). The conclusions that can be drawn from Fig. 5 and Fig. 6 are twofold. On the one hand, non-congested traffic conditions allow to safely send more than two messages per second, decreasing the mean error position estimate from 11 metres with a fixed 500ms CAM message interarrival time down to 4.48 metres with 200ms interarrival time (i.e., 5 messages per second). On the other hand, traffic jams allow vehicles to transmit CAM messages at a lower rate while not having a substantial impact in the error position estimate since their speed is also lower. This can be appreciated considering the evolution of the error estimate if CAM messages interarrival time is set to 1000ms, where it is observed that the error is reduced from 22.9m under non-congested traffic conditions down to 11.5m when the road is densely occupied.

After exploring the performance of modifying the CAM interval rate in terms of overload alleviation and the impact on the position estimate, we proceed to evaluate the performance of mobility load balancing.

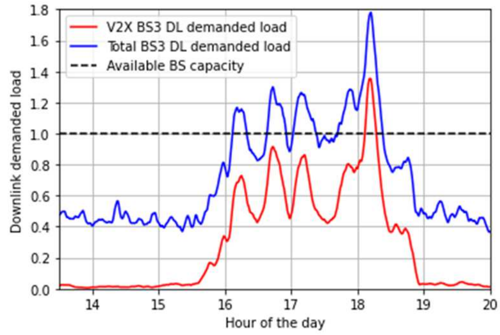

Fig. 3: Demanded traffic load at BS3 and its vehicular component.

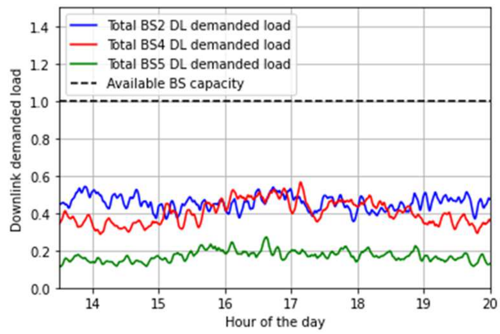

Fig. 4: Demanded load in neighbouring BS (i.e., BS2, BS4 and BS5).

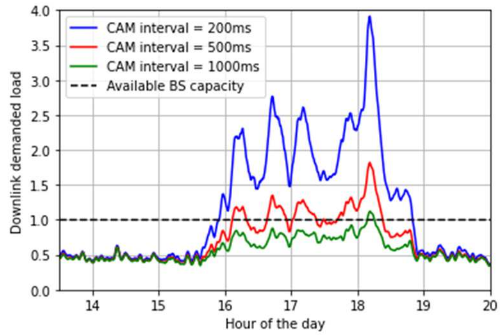

Fig. 5: Demanded load in BS3 depending on the CAM interarrival time.

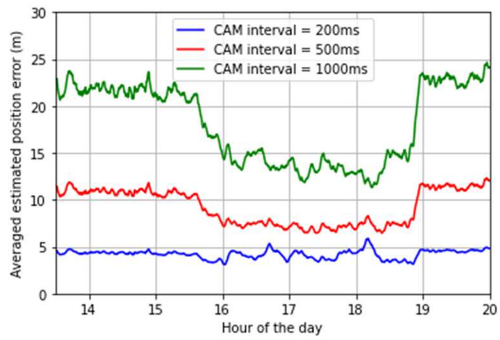

Fig. 6: Estimated position error in BS3 over time depending on the CAM interarrival time.

Enlighted by the preliminary results provided in [12], we focus on introducing a 10 dB handover offset between BS3 and BS4. Results are shown in Fig. 7 and Fig. 8, assuming a CAM interarrival time of 500ms. More precisely, Fig. 7 shows the effect load balancing has in BS3 if the offset is applied and Fig. 8 shows the transferred load to BS4. It is found that considering the timeframe from 16:00 until 19:00, if no shift had been applied the demand would have exceeded the available capacity during 53.14% of the time. In turn, if the offset is applied, the overload only occurs during 8.56% of the time (at around 18:00), which implies a reduction of the overload time in BS3 of 83.88%.

Having described the strengths and weaknesses of modifying the CAM message rate and the potentials of load balancing, we proceed to evaluate the performance of modifying the number of resources allocated to each slice. In this context, Fig. 9 and Fig. 10 exhibit the downlink load occupancy in BS3 from each slice with two resources configurations. We define the load occupancy as the portion of available resources that are being used by a given slice. In particular, Fig. 9 considers that the V2X resources are being sized according to their average demand and allocates 25% of the available resources to the vehicular slice. Thus, the remaining 75% of resources are allocated to the eMBB. In turn, the reverse strategy is applied in Fig. 10, trying to accommodate V2X resources according to their peak demand (75%) and assigning the remaining 25% to the eMBB slice. It can be observed that while the first approach ensures that eMBB slice gets the required resources throughout all the time frame (i.e. the eMBB load occupancy equals the eMBB demanded load during all the time), the resources allocated to the vehicular service fail to meet the required demand during all the congestion period, in which the load occupancy does not meet the demanded load. Instead, if the reverse strategy is applied, the service requirements are partially met during the overload compromising the eMBB slice.

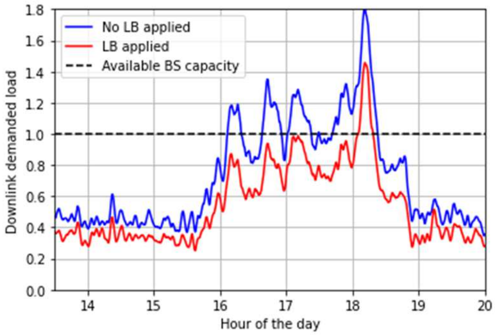

Fig. 7 Original demanded and balanced demanded load at BS3 with a fixed CAM interarrival time of 500ms.

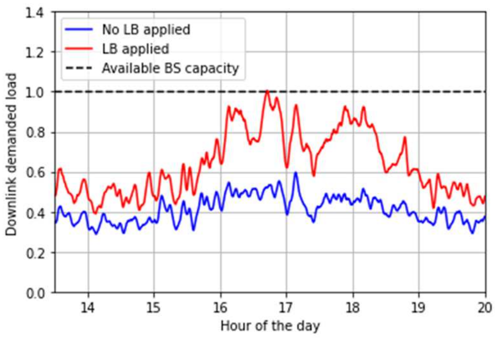

Fig. 8: Original demanded load at BS4 and transferred traffic load from BS3 with a fixed CAM interarrival time of 500ms.

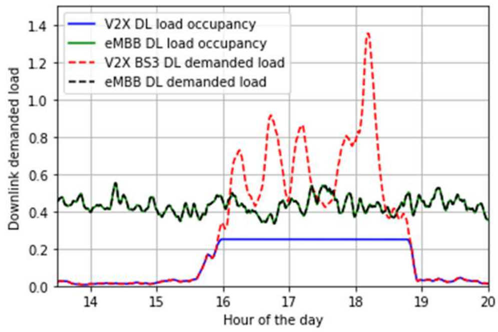

Fig. 9: Load occupancy for each slice in BS3 over time if 75% of resources are allocated to eMBB and 25% to V2X communications.

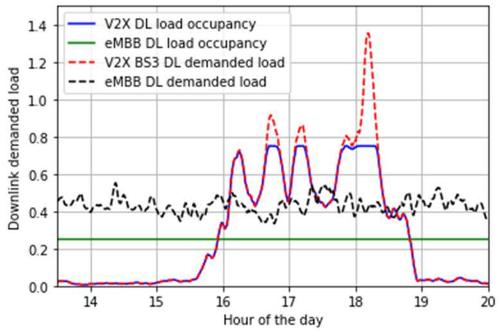

Fig. 10: Load occupancy for each slice in BS3 over time if 75% of resources are allocated to eMBB and 25% to V2X communications.

Considering the timeframe from 16:00 until 19:00, if only 25% of RB are allocated to the vehicular slice, this slice will only meet the demanded load during 6.94% of this timeframe. On the contrary, if allocating 75% of resources to the vehicular slice, the service will be satisfied in 70.70% of the time.

Keeping all the above results in mind and being the considered overload alleviation techniques complementary among them, some guidelines for an effective combination of these approaches can be provided. In this respect, the best method to reduce the total demanded load is decreasing the CAM message rate, provided that the application layer can support the associated service degradation. If this is not enough, diverting furthermost users to their neighbouring cells also alleviates the overload without degrading the performance of the services. In turn, modifying the assigned resources to each slice would only allow us to select which service would see its performance mostly degraded.

Thus, a comprehensive understanding of the vehicular services' requirements is of key importance to operate an efficient strategy combining the three techniques presented.

V. CONCLUSIONS AND FUTURE WORK

This paper has proposed and evaluated a set of different cell overload alleviation techniques in a vehicular network using realistic vehicular traffic traces and considering an urban scenario with V2X, eMBB and mMTC slices.

Starting from a cell overload introduced by a traffic congestion, modifying the CAM message generation rate has been proved to effectively reduce the network overload even compromising the quality of experience. Parallelly, using the information from vehicles 'mobility enables to introduce an efficient load balancing approach, which has helped to reduce the overload time in around 85%. Finally, the constraints from the way physical resources can be statically shared among different slices in a dynamic environment have been discussed, and the effect this has in the event of having a network overload detailed. Beyond the performance of the stand-alone evaluation of each strategy, a discussion on how they should be combined has been provided.

Based on this work, our future research envisages the algorithmic formulation to autonomously satisfy the dynamic requirements of vehicular slices in realistic scenarios. Through these algorithms, a dynamic implementation of the different alleviation techniques should be combined, enabling an efficient network operation in a network slicing framework.

ACKNOWLEDGEMENT

This research was supported by the Spanish Centre for the Development of Industrial Technology (CDTI) and the Ministry of Economy, Industry and Competitiveness under grant/project CER-20191015 / Open, Virtualized Technology Demonstrators for Smart Networks (Open-VERSO).

REFERENCES

[1] Global mobile suppliers association, "5G Network slicing for vertical industries," 2017.

[2] "5G Automotive Association," [Online]. Available: https://5gaa.org/. [Accessed 31 October 2021].

[3] 3GPP TS23.501, "5G, System architecture for the 5G System (5GS). (Release 16) v.16.6.0," Oct. 2020.

[4] X. Zhang, M. Peng, S. Yan and Y. Sun, "Deep-reinforcement-learning-based mode selection and resource allocation for cellular V2X communications," *IEEE Internet of Things Journal,* vol. 7, no. 7, pp. 6380-6391, 2020.

[5] P. Luoto, M. Bennis, P. Pirinen, S. Samarakoon, K. Horneman and M. Lavta-aho, "Vehicle clustering for improving enhanced LTE-V2X network performance," in *European Conference on Networks and Communications (EuCNC)*, Oulu, 2017.

[6] H. Khan, S. Smarakoon and M. Bennis, "Enhancing video streaming in vehicular networks via resource slicing," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 4, pp. 3513-3522, 2020.

[7] H. Wu, J. Chen, C. Zhou, J. Li and S. Xuemin, "Learning-Based joint resource slicing and scheduling in space-terrestrial integrated vehicular networks," *Journal of Communications and Information Networks,* vol. 6, no. 3, pp. 208-223, 2021.

[8] H. Daami R. Albonda and J. Pérez-Romero, "Analysis of RAN slicing for Cellular V2X and mobile broadband services based on reinforcement learning," *EAI Endorsed Transactions on Wireless Spectrum,* vol. 4, 2020.

[9] S. Uppoor and M. Fiore, Large-scale urban vehicular mobility for networking research, Amsterdam: VNC, 2011.

[10] L. Codeca, R. Frank, S. Faye and T. Engel, "Luxembourg SUMO Traffic (LuST) scenario: traffic demand evaluation," *IEEE Intelligent Transportation Systems Magazine,* vol. 9, no. 2, pp. 52-63, 2017.

[11] S. Uppoor and M. Fiore, "Characterizing pervasive vehicular access to the cellular RAN infrastructure: an urban case study," *IEEE Transactions on Vehicular Technology,* vol. 64, no. 6, pp. 2603-2613, 2015.

[12] M. Trullenque, O. Sallent, D. Camps, J. Escrig and C. Herranz, "Analysis of Vehicular Scenarios and Mitigation of Cell Overload due to Traffic Congestions," in *IEEE 2022 Spring Vehicular Technology Conference*, Helsinki, 2022.

[13] 3GPP TS 23.285, "Architecture enhancements for V2X services. (Release 16) v.16.3.0," Jul. 2020.

[14] 3GPP TS 23.286, "Application layer support for Vehicle-to-Everything (V2X) services; Functional architecture and information flows. (Release 16). v.16.6.0," Jul. 2021.

[15] ETSI TS102 894-1, *Intelligent Transport Systems (ITS) Users and applications requirements. Part 1: facility layer structure, functional requirements and specifications.,* Aug. 2013.

[16] ETSI TS 102 637 - 2, "Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications.," Mar. 2011.

[17] O. Shagdar, F. Nashashibi and S. Tohme, "Performance study of CAM over IEEE 802.11p for cooperative adaptive cruise control," in *2017 Wireless Days*, Porto, Portugal, 2017.

[18] ETSI, TR 102 698 Intelligent Transport Systems (ITS) Vehicular Communications C2C-CC Demostrator 2008 Use Cases and Technical Specificatoins, 2010.

[19] ETSI, EN 302 665 Intelligent Transport Systems(ITS); Communcatoins Architecture, 2010.

[20] ETSI, EN 302 663 (ITS) ITS-G5 Access layer specification for Intelligent Transport Systems operating in the 5 GHz frequency band, 2019.

[21] ETSI, Intelligent Transport Systems (ITS); Decentralized Congestion Control Mechanisms for Intelligent Transport Systems operating in the 5 GHz range; Access Layer part.

[22] A. Mansouri, V. Martinez and J. Härri, "A First Investigation of Congestion Control for LTE-V2X Mode 4," in *2019 15th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, Wengen, 2019.

[23] M. Roshdi, S. Bhadauria, K. Hassan and G. Fischer, "Deep Reinforcement Learning based Congestion Control for V2X Communication,," in *2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Helsinki, 2021.

[24] ETSI, TS 103 175 Intelligent Transport Systems (ITS); Cross Layer DCC Management Entity for operation in the ITS G5A and ITS G5B medium, 2015.

[25] a. Roshdy, A. Gaber, F. Hantera and M. ElSebai, "Mobility load balancing using machine learning with case study in live network," in *2018 International Conference on Innovative Trends in Computer Engineering (ITCE), 2018*, Aswan, 2018.

[26] P. Muñoz, R. Barco and d. l. B. I., "Optimization of load balancing using fuzzy Q-Learning for next generation wireless networks,," *Expert Systems with Applications,* vol. 40, no. 4, pp. 984-994, 2013.

[27] 3GPP TR 36.902, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions. (Release 9) v.9.3.1," May, 2011.

[28] CellMapper, "Signal tiles and towers," CellMapper, [Online]. Available: https://www.cellmapper.net/. [Accessed 4th September 2021].

[29] OpenStreetMap contributors, *Planet dump retrieved from https://planet.osm.org,* https://www.openstreetmap.org, 2017.

[30] 3GPP TS 38.104, "5G; NR; Base Station (BS) radio transmission and reception," Jul. 2018.

[31] 3GPP TR 36.942, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios. (Release 16) v16.0.0.0," Jul. 2020.

[32] 3GPP TS 38.214, "5G; NR; Physical layer procedures for data. (Release 15) v.15.3.0," Oct. 2018.

[33] 5GAA, "5GAA TR A-200094 Working Group System Architecture and Solution Development V2X Application Layer Reference Architecture," München, 2020.

[34] 3GPP TR 38.913, "Study on scenarios and requirements for next generation. (Release 14). v.14.2.0," May, 2017.

[35] L. M. W. Lopez, C. F. U. Mendoza, J. Casademont Serra and D. Camps Mur, "Understanding the impact of the PC5 resource grid design on the capacity and efficiency of LTE-V2X in vehicular networks," *Wireless communications and mobile computing,* vol. 2020, pp. 1-13, 2020.