AI-empowered management and orchestration of vehicular systems in the beyond 5G era

Reference:
Slamnik-Krijestorac Nina, Camelo Miguel, Chang Chia-Yu, Soto-Arenas Paola Andrea, Cominardi Luca, De Vleeschauwer Danny, Latré Steven, Marquez-Barja Johann.- AI-empowered management and orchestration of vehicular systems in the beyond 5G era
IEEE network / IEEE [New York, N.Y.] - ISSN 1558-156X - 37:4(2023), p. 305-313
Full text (Publisher's DOI): https://doi.org/10.1109/MNET.008.2300024
To cite this reference: https://hdl.handle.net/10067/2014380151162165141

AI-empowered Management and Orchestration of Vehicular Systems in the Beyond 5G era

Nina Slamnik-Kriještorac*, Miguel Camelo*, Chia-Yu Chang†, Paola Soto*, Luca Cominardi‡, Danny De Vleeschauwer†, Steven Latré*, and Johann M. Marquez-Barja*
* University of Antwerp - imec, IDLab - Faculty of Applied Engineering, Belgium
† Nokia Bell Labs, Belgium
‡ Zettascale, France

***Abstract*—The complexity of orchestrating Beyond 5G services, such as vehicular, demands novel approaches to remove limitations of existing techniques, as these might cause a large delay in orchestration operations, and thus, negatively impact the service performance. For instance, the human-in-the-loop approach is slow and prone to errors, and closed-loop control using rule-based algorithms is difficult to design, as an abundant number of parameters need to be configured. Applying Artificial Intelligence (AI)/Machine Learning (ML), in combination with Network Function Virtualization (NFV) and Software Defined Networking (SDN), seems a promising solution for enabling automation and intelligence that will optimize orchestration operations. In this paper, we i) study the challenges in current ETSI NFV orchestration solutions for B5G C-V2X edge services, ii) propose an AI/ML-based closed-loop orchestration framework, iii) propose how and which AI/ML techniques can alleviate the identified challenges and what are the implications resulting from applying certain AI/ML techniques, and iv) discuss AI/ML-based system enablers for B5G C-V2X services.**

***Index Terms*—management, orchestration, AI/ML, C-V2X, MEC, NFV, SDN, NIF**

I. Introduction

The Beyond 5G (B5G) ecosystem will offer extreme-low latency (below 1 ms), ultra-high reliability (99,999999%), and enhanced throughput (user-experienced rate up to 1 Gbps), with applications deployed at the network edge as per European Telecommunications Standards Institute (ETSI) Multi-Access Edge Computing (MEC) framework [1]. As such, B5G is creating opportunities for verticals to improve services and create new ones that were not feasible before [2], such as those shown in Fig. 1: i) maneuver recommendation that instructs vehicles on the path/speed, ii) collision avoidance, iii) teleoperation, and iv) infotainment (e.g., video streaming). To realize such services, B5G radio access and core networks are used together with the orchestrated Network Function Virtualization (NFV) infrastructure [3], spanning different edge domains (Fig. 1).

To provide Cellular Vehicle-to-Everything (C-V2X) services in a reliable manner by localizing access to virtualized network resources, challenges such as operating under constrained resources, with heterogeneous network edges, and in time-varying conditions, must be carefully addressed [3]. These challenges are particularly important in highly mobile environments with connected vehicles, where MEC-based C-V2X services require continuous monitoring of network/computing resources, and efficient service control (e.g., fast scaling, re-deployment, migration) [3]. Therefore, the traditional manual network management becomes impossible to scale and to maintain, as it is either open-loop and inherently manual or closed-loop but slow (e.g., human-in-the-loop) and based on simple and static rules. In addition, B5G systems are expected to connect a massive number of heterogeneous devices connecting to the network over a diverse set of network slices stretching over numerous technological domains (radio, edge,

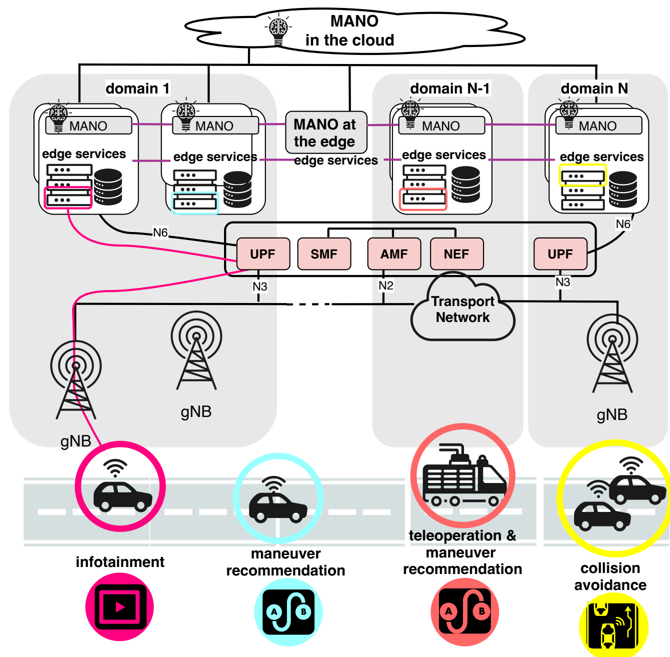


Fig. 1: NFV MANO in B5G C-V2X system with edge services deployed in different edge network domains.

core), which brings severe challenges to traditional centralized management and orchestration approaches towards achieving high scalability and sustainability [4].

To address the aforementioned challenges, pervasive Artificial Intelligence (AI)/Machine Learning (ML) techniques are recognized as key enablers for increasing automation and efficiency of Management and Orchestration (MANO) operations [5]. This way, service placement/scaling/migration can be designed to achieve service continuity by leveraging data analytics and AI/ML techniques for event anticipation, fast response, and in-advance preparation of the network. For instance, the ETSI NFV considers the incorporation of AI/ML into their standardized MANO stack for MEC services [1], but only on the level of rule-based auto-scaling/auto-healing policies defined through the use of service descriptors. As these descriptors only specify how the scale or heal actions need to be executed when a condition is satisfied, the full automation procedures are not yet tackled by ETSI NFV, but by ETSI Zero-touch Service Management (ZSM) and Experiential Networked Intelligence (ENI). ENI is considered as a closed-loop policy-driven AI/ML that combines existing technologies (data analytics and AI mechanisms), and ZSM is specifying solutions for end-to-end service orchestration including enablers for closed-loop automation. However, neither of them is considering vertical-specific requirements such as those in the C-V2X environment, and challenges that arise from applying AI/ML mechanisms for C-V2X service orchestration, where such generalization could significantly hinder service operation.

Despite the presence of many open-source MANO tools, the support for an agile and automated service re-configuration still remains challenging due to the complexity of orchestrated environments where applying AI/ML is not a straightforward task. Thus, there is still a lack of full understanding of the selection, deployment, and impact of AI/ML on the NFV MANO operations for C-V2X services. To this end, this article:

- identifies the challenges in current NFV MANO solutions for B5G-based C-V2X services, analyzing the AI/ML potential to mitigate those challenges, going beyond the state-of-the-art of standardization bodies [6],
- studies on how/which AI/ML techniques have the potential to improve the MANO operations,
- analyzes the Network Intelligence (NI) system enablers for C-V2X services, and
- identifies implications of applying AI/ML that need to be further studied.

Concerning related work, Chergui et al. [4] propose a decentralized 6G zero-touch management framework for improving management scalability issues, and for reducing reaction times of self-configuration and self-healing processes. In this case, scalability is achieved by penetrating decision-making power into distributed network elements performing life-cycle management of network slices, and zero-touch management by enabling a closed-control loop. In addition, Grasso et al. [7] propose a Deep Reinforcement Learning (DRL)-based network management framework for Unmanned Vehicle (UAV) edge networks, showcasing significant improvements in the task execution time on distributed edge computing nodes hosting delay-sensitive applications.

Compared to the other works [4,5,7–9] that present the general concept of applying AI/ML to NFV MANO, this work is, to the best of our knowledge, among the first ones that provide such study for C-V2X service orchestration.

II. Toward Automated MANO for V2X

A. Challenges in the current NFV MANO solutions

The NFV MANO systems perform service placement, scaling, migration, and termination, based on information gathered from various network segments. By studying the existing solutions and their applicability to C-V2X service orchestration, we identify several challenges that need to be carefully addressed.

a) Manual orchestration operations: The stringent requirements for C-V2X services, with self-driving vehicles as an ultimate goal, require extensive (uplink) broadband and reliable connectivity of up to five-nines [2]. This urges for real-time monitoring of the network performance to achieve improved decision-making.

b) Insufficient operational efficiency: of NFV MANO operations needs to be improved (e.g., lengthy scaling procedure that hinders service reliability and response time), as i) processing monitored data and making decisions are traditionally manual and require human intervention that is prone to mistakes and additional delays, and ii) network complexity significantly increases with heterogeneous and distributed resources [10], which is even more significant in C-V2X systems because of the presence of various automobile manufacturers, vehicle application providers, and Mobile Network Operators (MNOs). The application of AI/ML to enable automated NFV MANO operations combines data analytics and learning in closed-loop, thereby outperforming complex and lengthy optimization schemes, heuristic ones that are problem-specific and domain-dependent, and open-loop approaches that are prone to human errors, which makes them all ineffective in swift responses to dynamic network changes and vehicle mobility [10].

c) KPI fluctuations: occur due to fluctuations in the demands from vehicles, and their mobility patterns, which is particularly challenging when large numbers of moving vehicles are simultaneously connected to the orchestrated edge services. Thus, orchestrators need to improve their operation by learning from the environment, identifying or even predicting changes in Key Performance Indicators (KPIs), and translating these changes into required NFV MANO operations that will maintain service performance at the desired level.

d) Increased load of NFV MANO: From 5G onward, the virtualization is realized in the core network, and partially on the radio side. Also, vehicles are getting equipped with computing units, which become mobile edge nodes that can host Virtualized Network Functions (VNFs). Therefore, NFV MANO solutions are expected to orchestrate all these VNFs. Such an ever-increasing load on the MANO solution may hinder the performance of its operations within the response time required to capture fluctuating KPIs. This phenomenon can be detrimental to C-V2X performance (e.g., increased response time from a C-V2X edge service to the vehicle due to insufficient computing/network resources) and must be prevented.

e) Insufficient and inconsistent input data: Huge amounts of data are collected from surrounding infrastructure (edge nodes, sensors, gNodeBs) for orchestrators to coordinate distributed service deployments, which is more complex than in centralized clouds. This becomes more challenging due to the mobility and varying network connectivity, which may cause delays or jitters in data collection. This lack of sufficient and consistent input data leads to inefficiencies in decision-making, e.g., where/when to migrate service from one edge to another.

f) Support for multi-domain orchestration: The access to C-V2X edge services should be ensured across different domains, as vehicles move along the roads, traversing from one edge domain to another. To this end, coordination among multiple orchestrators is required. Such MANO operations are performed across different NFV domains for particular C-V2X services (e.g., services that send maneuver recommendations to vehicles in more than one domain to avoid road congestion), and can be realized by using particular learning framework.

B. The Need for Automated and Intelligent MANO for V2X

Addressing the previous challenges will transform traditional MANO for C-V2X systems into a fully autonomic system able to adapt the services and infrastructure to changes in user demands, business goals, and/or environmental conditions. In particular, AI/ML could provide the NI for MANO systems through the Network Intelligence Functions (NIFs), which are the pipelines of effective AI/ML algorithms that detect/anticipate new requests or fluctuations in the network activities [11], and help orchestrators to respond to such changes in a fully automated manner.

Unlike the legacy analytical-based models with many parameters that can affect KPIs, the data-based models are enabling a closed-loop approach to perform MANO of C-V2X services (Fig. 2), which is crucial for automation and optimization. This is precisely where AI/ML will play a fundamental role.

To mitigate the challenges listed in Section II-A, we propose to integrate AI/ML techniques into a closed-loop framework to

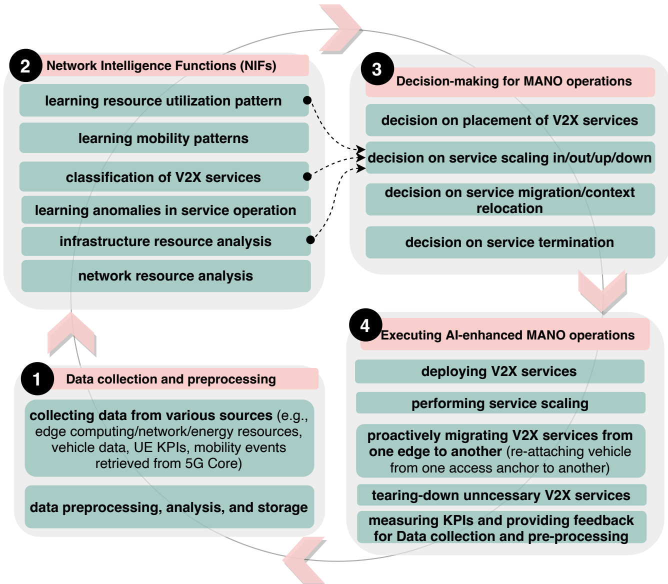


Fig. 2: Closed-loop framework for NFV MANO in C-V2X systems; The dashed arrows showcase an example on how the decision on service scaling can be made based on the three different NIFs.

realize a fully autonomous NFV MANO, as these techniques are now sufficiently mature to provide solutions for complex optimization and decision-making processes.

However, due to the complexity and heterogeneity of C-V2X systems, it is impractical to automate MANO operations by applying a single ML model per MANO operation. On the contrary, suitable AI/ML techniques should be applied as NIFs, which focus on a particular task (e.g., mobility pattern, resource utilization), whose outcomes are then jointly considered in NFV MANO, where the final orchestration decision is made. Thus, in Fig. 2, we define the following phases of a closed-loop framework:

- *Data collection and pre-processing*: in charge of collecting data from various sources, which is then pre-processed and shared with the NIFs that apply corresponding AI/ML techniques.
- *NIFs*: get the relevant data, and make predictions/decisions that support orchestrators towards improving their operations.
- *Decision-making for MANO operations*: instantiation/scaling/migration/termination is performed based on the decisions that are harmonizing outputs from a group of NIFs. For example, in Fig. 2, we showcase how the decision on service scaling should be made considering the outputs from learning resource utilization pattern, infrastructure resource analysis, and further adjusting the decision to a particular service class that is identified by the NIF that classifies C-V2X services.
- *Executing AI-enhanced MANO operations*: usually performed by edge platform and virtualized infrastructure managers, which apply decisions made by orchestrators, and re-configure service deployments.

The closed-loop framework we propose is generic, but some widely used frameworks described in [12], such as Monitor-Analyze-Plan-Execute-Knowledge (MAPE-K), and Observe-Orient-Decide-Act (OODA), can be applied.

III. AI/ML SOLUTIONS FOR NFV MANO OPTIMIZATION AND AUTOMATION

To provide tangible ideas for the phases described in Section II-B, we elaborate on the six examples of NIFs listed in Table I. Table I also shows the data to be collected for the corresponding NIF, the potential AI/ML technique to implement it, and the list of the relevant C-V2X service types (Fig. 1) that would be impacted by them. The final decision by the NFV orchestrator is then applied by selecting outputs of different NIFs via simpler approaches, such as rule-based and multi-criteria decision-making frameworks, resulting in various resource re-configurations (e.g., resource reservation/release). The life-cycle management of NIFs and ensuring their proper functioning needs to be performed by the NI orchestration layer. This orchestrator is out of the scope of this paper, but in Section V, we provide a glimpse of its role and importance.

Following, we will introduce each of the proposed NIFs (Phase 2 in Fig. 2), and elaborate on the potential candidate AI/ML techniques to be applied, based on the analysis of eight well-known AI/ML techniques that are presented in Fig. 3 together with the challenges identified in Section II-A that these techniques are expected to alleviate.

a) Learning resource utilization patterns for different service types: aims at learning the resource utilization patterns to i) determine resource requirements by looking for spatiotemporal correlations in historical data (Table I), and ii) enable resource elasticity through forecasting resource utilization, which is important given the resource constraints at the edges. Based on the analysis shown in Fig. 3, Supervised Learning (SL) can use the labeled historical data to determine the relationship between edge computing and network resources that are provisioned to the service, and the KPIs measured at the client. Given this relationship, KPIs are determined based on the predicted resource utilization (e.g., regression models), and thresholds are defined to provide a finer-grained estimate of KPIs for a particular service when deployed at edge. Such a model can support the orchestrators in making decisions on service instantiation (e.g., edge node selection), or proactive scaling for maneuver recommendation and teleoperation services, as they are critical when it comes to a service response time that needs to be monitored.

b) Learning mobility patterns: by applying SL/Unsupervised Learning (UL) is beneficial for edge node selection. This task needs to consider the data collected from the vehicles, informing NIF about the speed, location, and heading of all connected vehicles, including data from the B5G core, e.g., mobility event notifications from Access and Mobility Management Function (AMF) [13], and looking for the spatiotemporal correlations of the vehicles' locations. In addition to considering mobility patterns when optimally placing infotainment services, the decisions made in this NIF are provided as input for the orchestrator to deciding about C-V2X service migration. This is particularly important for collision avoidance services as they can be migrated to the edges closer to a dense group of vehicles that need to prevent collisions.

c) Classification of V2X services: aims to select the network slices for each service, thereby improving the network Quality of Service (QoS) and Quality of Experience (QoE) perceived by the vehicles and ensuring compliance with the Service Level Agreement (SLA). Therefore, SL can be used to classify on-demand C-V2X services (e.g., infotainment and teleoperation) and services that are always deployed on edge nodes (e.g., maneuver recommendation and collision avoidance). This task can also assign priorities to be considered when deciding which services to teardown/mute, and which ones to scale up to improve their performance.

d) Learning anomalies in service operation: is expected to use Online Learning (OL) to identify run-time anomalies. Taking the infotainment service as an example, an offline-trained ML model may not respond effectively when there

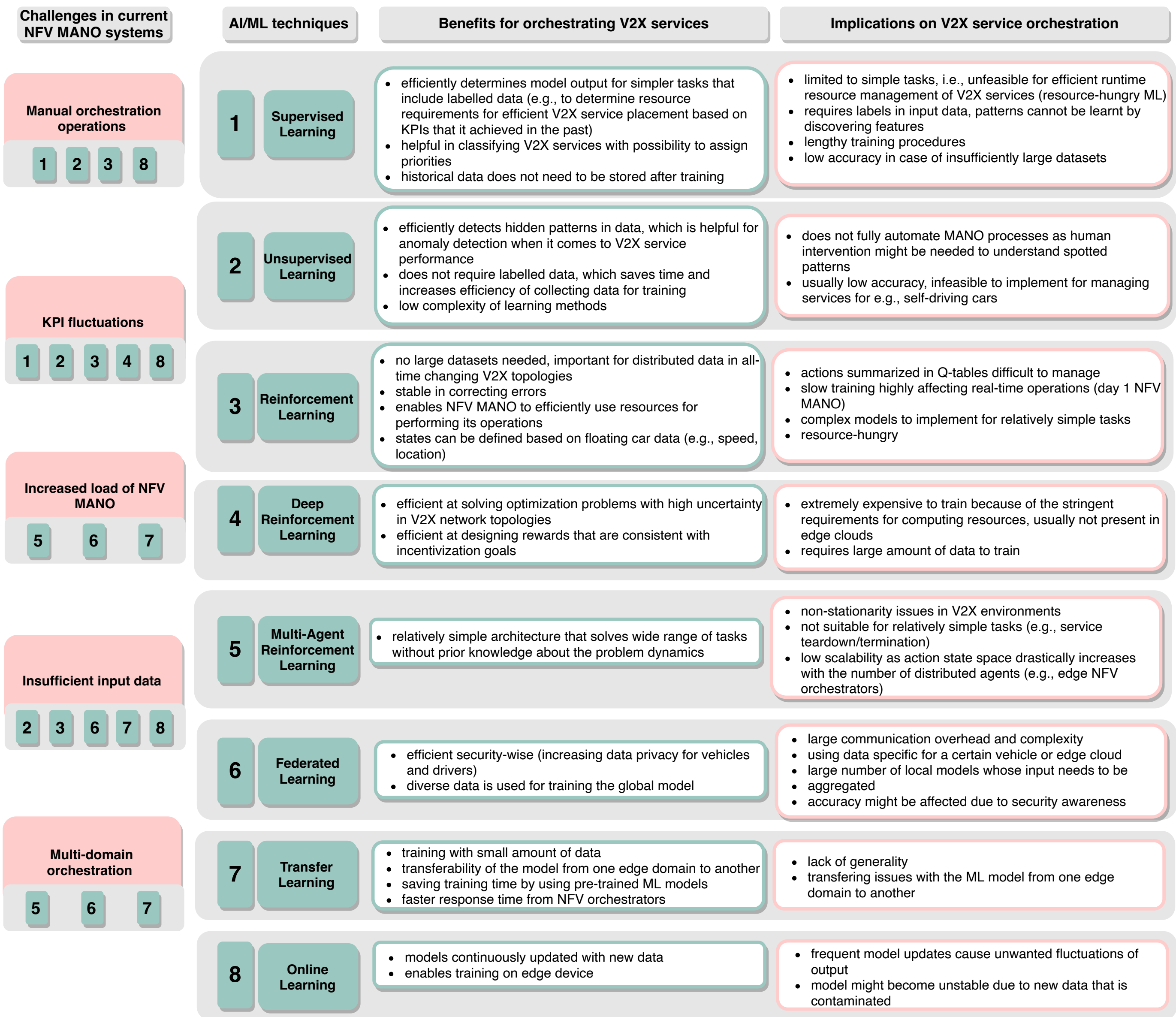


Fig. 3: Overview of challenges in current NFV MANO solutions for C-V2X services, and potential solutions in the form of the most commonly used AI/ML models.

is a surge in the number of vehicles playing a specific live video stream. Scaling operations are not triggered correctly, resulting in a decline in the QoE. Another example is the decision-making model for scaling/migrating collision avoidance services, in which new data streams must be updated because vehicle collisions can occur sporadically. In the case of OL, models can learn in seconds and minutes, and update themselves based on new input data. This makes OL suitable for such NIF in C-V2X systems (e.g., data in motion), where new data streams from moving vehicles are constantly generated.

e) Edge infrastructure and network resource analysis: performs a complex global optimization task given the ever-changing C-V2X topology and traffic fluctuations. Once thoroughly trained, Reinforcement Learning (RL) is robust and stable; thus, it is a promising technique for this NIF, especially because it uses an intelligent agent to learn by interacting with the environment in a closed-loop manner. Nevertheless, due to the high resource requirements, the application of RL should be carefully considered and used only to address large-scale optimization of resources across multiple edge nodes. One solution is to use RL on the cloud-level orchestrators, where such a model can determine the etwork traffic state, e.g., determining the congestion zones for maneuver recommendation and infotainment services, and to help redirect vehicles, e.g., publishing recommendations/statistical analysis of congestion on the roads, in a common repository available to all edge orchestrators. Therefore, service placement/migration can be realized without supporting RL technique at each edge. For cooperative maneuvering of (automated) vehicles, a collaboration between edge orchestrators is needed, and Multi-Agent Reinforcement Learning (MARL) can deploy multiple learning agents at different collaborative edges. An important benefit of MARL is its relatively simple architecture that solves a wide range of tasks without prior knowledge of the problem dynamics. This is important for continuously changing environments with resource fluctuations. However, if the cloud orchestrator deploys a C-V2X service on a particular

TABLE I: Mapping the identified gaps to the proposed NIFs in the closed-loop framework for NFV MANO.

Data collection and pre-processing	**Network Intelligence Functions (NIFs)**			**** MANO operations**	**V2X service type**
	Role	**Description**	*** ML**		
edge computing resources, network resources, KPIs measured at client side	learning resource utilization patterns for different types of services	looking for spatio-temporal correlations in historical data, and forecasting resource utilization	SL, UL	I, M, S	infotainment, teleoperation, maneuver rec.
mobility events from Core network, floating car data (speed, location, heading)	learning mobility patterns	reducing dimensionality of multiple source information; finding spatio-temporal correlations of the vehicles' locations; scheduling V2X services to the neighboring edge nodes by precaching relevant content and balancing the load	SL, UL, FL	M	maneuver rec., collision avoidance
edge computing resources, network resources, KPIs measured at client side	classification of V2X services	mapping between QoS metrics and service priorities; assigning V2X service to a priority/non-priority slice	SL, UL	I, S, M, T	infotainment, teleoperation, maneuver rec., collision avoidance
KPIs measured at client side	learning anomalies in service operation	filtering anomalies as deviations from normal behavior; identifying the reckless driving maneuvers; isolating anomaly by allocating less resources to attacking sources	SL, UL, OL	S, M	infotainment, collision avoidance
edge computing resources, energy consumption	infrastructure resource analysis	evaluating computing resources /energy consumption trends for next operation hours; scheduling turning on/off the critical edge nodes according to computing resources/energy consumption plans	SL, FL, RL, MARL	I, M	infotainment, maneuver rec.
network resources (bandwidth, latency), mobility events from Core network	network resource analysis	predicting QoS metrics from current network state	SL, FL, RL, MARL	I, M	infotainment, maneuver rec.

* Supervised Learning (SL), Unsupervised Learning (UL), Federated Learning (FL), Online Learning (OL), Reinforcement Learning (RL), Multi-Agent Reinforcement Learning (MARL)

** Instantiation (I), Scaling (S), Migration (M), Termination (T)

edge that lacks collected data from the network infrastructure, Federated Learning (FL) can apply a global ML model, which is evaluated in local edge domains based on their data, to help predict resource usage on this particular edge node.

IV. Network Intelligence in V2X ecosystem

In addition to the above introduced NIFs for MANO in C-V2X systems, we further examine the essential elements for implementing an overall NI system.

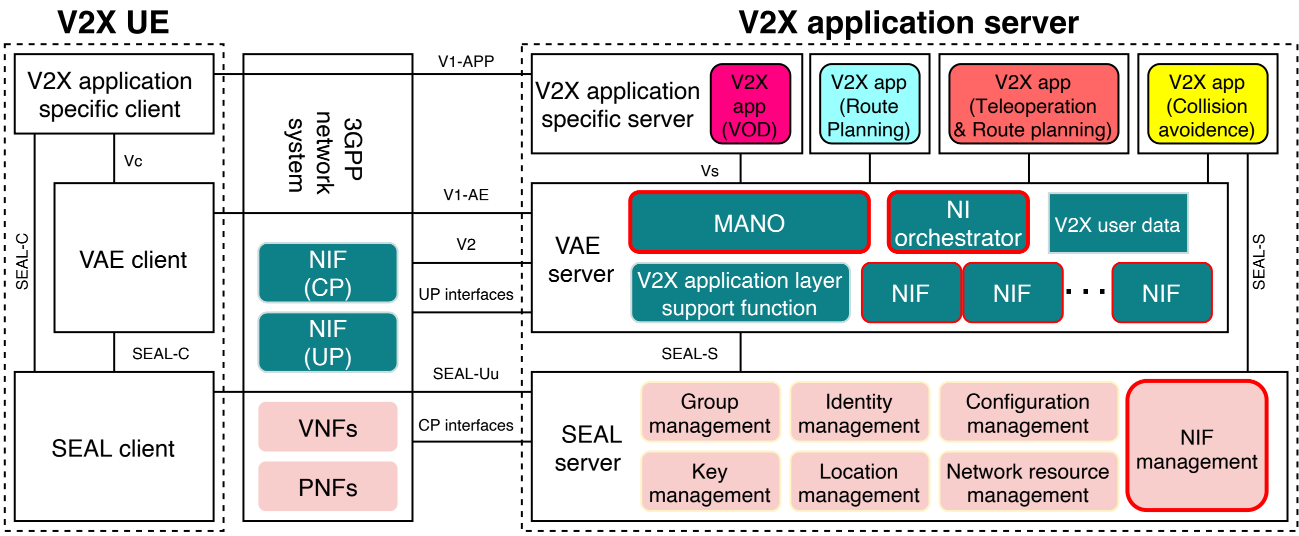
To introduce how NIFs fit the C-V2X ecosystem, Fig. 5 illustrates the relationship between different providers in the value chain transformation of the C-V2X industry, in line with 3^{rd} Generation Partnership Project (3GPP) TS23.286 [14]. First, providers of network infrastructure, network functions, communication services, and C-V2X services, are decoupled from C-V2X service users to allow cost-effective and flexible C-V2X service composition. Additionally, a new role is expected to provide NIFs in the form of AI/ML algorithms outlined in Section III. Specifically, these NIFs should not only be supported at MANO layer but also for the corresponding network slice(s) in Control Plane (CP) and User Plane (UP). Taking the second gap in Section II-A as an example, deploying the NIFs for MANO operations (e.g., scaling) may not be sufficient to respond promptly to KPI fluctuations, and C-V2X service users can notice performance degradation. To this end, NIFs in CP and UP can play a key role in adjusting scheduler policies and manipulating packets (e.g., packet marking/dropping), respectively.

Moreover, we propose an enhancement for the C-V2X application layer functional model from 3GPP, to be able to manage NIFs for C-V2X services and corresponding network slices, as shown in Fig. 4. We can see that the C-V2X application layer support functions at the V2X Application Enabler (VAE) layer exploit several Service Enabler Architecture Layer (SEAL) services to support C-V2X applications operations (see 3GPP TS23.286 and TS29.486). We propose three additional functional entities. First, the NIF management service at the SEAL server, which can be exploited by means of SEAL server interfaces to interact with the 3GPP network system for modifying NIFs. Second, NI orchestrator at the VAE server, provides support functions to communicate the requested NIFs to the 3GPP network and manages the applied NIFs. This NI orchestrator can harmonize different NIFs running inside the C-V2X application. Third, the MANO entity at the VAE server, which combines input from different NIFs, enabling C-V2X applications to interact with the overall MANO system and impact the effectiveness of their decision-making process.

V. Open Challenges

As a crucial step towards leveraging AI/ML to inject NI in network management, we provided an in-depth understanding of the true impact of AI/ML on the NFV orchestration operations.However, dealing with aftermath of applying AI/ML in dynamic C-V2X environments is not straightforward either, as AI/ML techniques impose additional challenges that shall be carefully considered.

a) Quality of data: The performance of AI/ML in decision-making depends on how close the training data is to the actual data used in the production environment. The lack of real-world samples may impose unmeasurable risks when training models are based on synthetic data (risks for drivers' safety). However, collecting data is time-consuming,



Functional entity & Reference point	Descriptions
V1-APP & V1-AE	V1-APP supports interaction related to V2X applications, V1-AE supports interaction related to V2X application layer support functions.
V2	Between VAE Server and V2X control function. The V2X control function provides UE parameter for V2X communication.
Vs, Vc	Support interactions related to V2X application layer support function and V2X application specific server/client.
SEAL-S, SEAL-C, SEAL-Uu	Include a number of reference points for respective management in SEAL architecture.
V2X application specific server/client	Provide server/client side functionalities (e.g., platooning server) corresponding to the V2X applications.
VAE server/client	V2X Application Enabler (VAE) server/client provides the client/server side V2X application layer support functions.
SEAL server/client	Various Service Enabler Architecture Layer (SEAL) services offer certain functionalities as APIs for vertical applications.

Fig. 4: The overview of V2X application layer functional model with NIFs, NFV MANO, and NI orchestrator, which is based on the framework for the application layer support for C-V2X services [14].

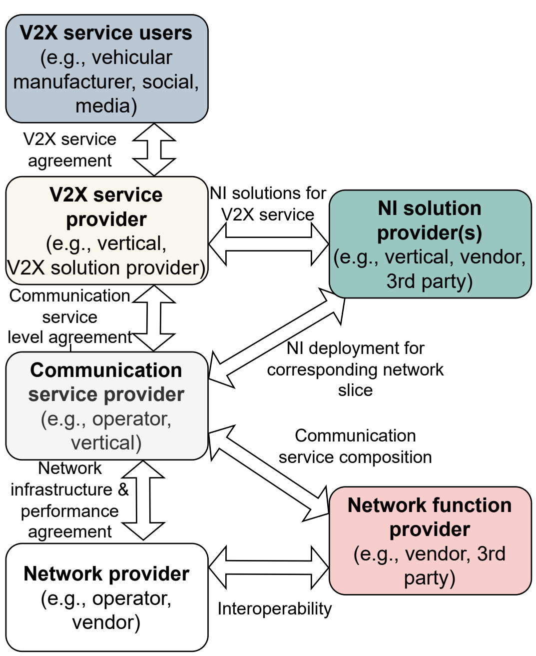


Fig. 5: Different providers in the value chain for the V2X industry.

as scenarios that require specific MANO operations (e.g., service scaling during natural disasters) are difficult to replicate and repeat to collect sufficient data for (re)training.

b) Security, scalability, and transferability: The AI/ML solutions are only as reliable as the data upon which they are trained. This is especially important as some C-V2X services need to assist their users through potentially life-threatening situations (e.g., collisions, and connectivity lost during teleoperation). In terms of security, particularly for vehicle data (e.g., vehicle identification, location/destination/speed), one possible solution is to apply an advanced identity and access management framework where vehicles are authenticated, authorized, and represented by security tokens that are stored only on specific edge servers. Regarding scalability and non-stationarity issues in RL and MARL, the former is due to the drastic increase in the action state space as the number of agents increases, and the latter is due to the decisions being influenced by the actions of other agents. An example is the relocation of emergency services based on the action taken by the source edge orchestrator, whereas the target edge orchestrator decides to completely mute all other services due to existing prioritized maneuvering operations. This situation must be prevented to avoid conflicting decisions.

c) Resource constraints: As C-V2X services are running concurrently with other edge services, this will affect the overall resource availability at the edge. However, resource-constrained edges may not be able to offer high computational power, making them unsuitable for running heavy data-processing tools (e.g., Apache Spark and deep learning libraries). The imbalance between lightweight implementation and high performance requires further study. Some efforts have been made to reduce computational and memory loads by applying model compressions, such as model pruning, parameter quantization, low-rank decomposition, knowledge distillation, and lightweight model design [15]. In the future, NI orchestrators should consider adapting models trained in the cloud before deploying them at the edge, with knowledge of the different trade-offs to make (e.g., model size vs. accuracy vs. energy consumption vs. inference time).

d) Proactive fault tolerance: Despite the prediction capabilities of intelligent MANO, their accuracy remains challenging. Since every pattern has exceptions (data outliers), the proactive MANO operations may make incorrect decisions when such prediction errors occur. Thus, it is necessary to study the extent to which predictors may make mistakes and determine whether they have serious consequences for service performance and whether plan B (e.g., reactive approach) should be prepared. This might not be an issue for infotainment and maneuver recommendation services, as their demands can be tested from a large number of vehicles (e.g., video content or route notifications). However, teleoperation and collision avoidance services require careful planning and preparation for data collection and testing. One possible approach to address this challenge is to use digital twins that mimic the real environment so that algorithms can be trained in a safe environment, but close enough to the one where they are deployed.

e) Update frequency of ML models: Although OL can be applied to learn anomalies during service operation, too frequent model updates may cause unwanted output fluctuations. However, if the update frequency is too low, catastrophic forgetting can occur, where previously learned knowledge is forgotten due to non-stationary data. Therefore, we expect future NI orchestrators to continuously monitor model performance, thereby adjusting the frequency of model updates based on the vehicle location and mobility patterns.

f) Modular design of B5G services: It is important to properly modularize the overall MEC service, either C-V2X or NIF, in a set of loosely coupled applications that can be migrated/scaled according to the decisions made by MANO or NI orchestrators. To support dynamic C-V2X environments, applications should rely on middlewares providing location-transparent communication and data access (e.g., Zenoh[1]) that is not hindered by the ever-changing underlying network topology and infrastructure. This allows critical applications and MANO to fall back to default safe mode, e.g., when enhanced C-V2X service and/or NIF at the edge suffers from unpredictable performance, autonomous vehicles may slow down while the MANO migrates/scales the involved applications.

g) NI Orchestration Layer: NIFs empowered by AI/ML require different life-cycle management compared to edge C-V2X services (e.g., model training, loss function adaptation, and resource-awareness). To fully support a complex, pervasive, and distributed nature of NI, an NI Orchestration layer should be introduced to manage intelligence as a whole, ensuring the ideal functioning of each closed-loop NIF, and overseeing interactions across closed loops that run NI at different timescales [11]. Thus, the NI orchestration layer must be carefully designed to synchronize the work of different NIFs. As B5G systems are expected to embed AI/ML in all network segments and functions (both in control and user planes), coordination of diverse NIFs and harmonization of their performance will be essential for leveraging the full potential of network intelligence and programmability. On the contrary, the efficient work service orchestrators (MANO systems) might be diminished if NIFs are producing inefficient decisions. This coordination of NIFs is also particularly important in multi-domain scenarios, where different orchestrators might collaborate toward improving their operations, but their base NIFs will also need to work in a coordinated manner to improve the quality of cross-domain decisions. In addition, the NI orchestration layer is also expected to perform NIF selection based on the NIF performance (e.g., inference delay) and C-V2X service requirements (expected end-to-end latency), as large inference times might prolong the orchestration decisions, and as such diminish service performance.

[1]Zenoh framework: https://zenoh.io/

ACKNOWLEDGEMENT

This work has been performed in the H2020 project DAEMON (Grant Agreement No. 101017109).

REFERENCES

[1] ETSI, "Multi-Access Edge Computing (MEC); Framework and Reference Architecture," *ETSI ISG MEC, ETSI GS MEC 003 V2.1.1*, 2019. Online [Available]: https://www.etsi.org/deliver/etsi_gs/MEC/001_099/003/02.01.01_60/gs_MEC003v020101p.pdf, Date accessed: 2023-05-01.

[2] H. Tataria, M. Shafi, A. F. Molisch, M. Dohler, H. Sjöland, and F. Tufvesson, "6G Wireless Systems: Vision, Requirements, Challenges, Insights, and Opportunities," *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021. doi: http://dx.doi.org/10.1109/JPROC.2021.3061701.

[3] N. Slamnik-Krijestorac, G. M. Yilma, M. Liebsch, F. Z. Yousaf, and J. Marquez-Barja, "Collaborative orchestration of multi-domain edges from a Connected, Cooperative and Automated Mobility (CCAM) perspective," *IEEE Transactions on Mobile Computing*, vol. 22, no. 4, pp. 2001–2020, 2023. doi: https://doi.org/10.1109/TMC.2021.3118058.

[4] H. Chergui, A. Ksentini, L. Blanco, and C. Verikoukis, "Toward Zero-Touch Management and Orchestration of Massive Deployment of Network Slices in 6G," *IEEE Wireless Communications*, vol. 29, no. 1, pp. 86–93, 2022. doi: http://dx.doi.org/10.1109/MWC.009.00366.

[5] E. Baccour, M. S. Allahham, A. Erbad, A. Mohamed, A. R. Hussein, and M. Hamdi, "Zero Touch Realization of Pervasive Artificial Intelligence as a Service in 6G Networks," *IEEE Communications Magazine*, vol. 61, no. 2, pp. 110–116, 2023. doi: http://dx.doi.org/10.1109/MCOM.001.2200508.

[6] ETSI, "Autonomous Networks, supporting tomorrow's ICT business," *ETSI White Paper No. 40*, no. 40, 2020. Online [Available]: https://www.etsi.org/images/files/ETSIWhitePapers/etsi-wp-40-Autonomous-networks.pdf, Date accessed: 2023-05-01.

[7] C. Grasso, R. Raftopoulos, and G. Schembra, "Smart Zero-Touch Management of UAV-Based Edge Network," *IEEE Transactions on Network and Service Management*, vol. 19, no. 4, pp. 4350–4368, 2022. doi: http://dx.doi.org/10.1109/TNSM.2022.3160858.

[8] S. Fu, F. Yang, and Y. Xiao, "AI Inspired Intelligent Resource Management in Future Wireless Network," *IEEE Access*, vol. 8, pp. 22425–22433, 2020. doi: http://dx.doi.org/10.1109/ACCESS.2020.2968554.

[9] M. Bagaa, T. Taleb, J. Riekki, and J. Song, "Collaborative Cross System AI: Toward 5G System and Beyond," *IEEE Network*, vol. 35, no. 4, pp. 286–294, 2021. doi: https://doi.org/10.1109/MNET.011.2000607.

[10] D. M. Manias and A. Shami, "The Need for Advanced Intelligence in NFV Management and Orchestration," *Netwrk. Mag. of Global Internetwkg.*, vol. 35, p. 365–371, Mar. 2021. doi: https://dx.doi.org/10.1109/MNET.011.2000373.

[11] M. Camelo, L. Cominardi, M. Gramaglia, M. Fiore, A. Garcia-Saavedra, L. Fuentes, D. D. Vleeschauwer, P. Soto, N. Slamnik-Kriještorac, J. Ballesteros, C.-Y. Chang, G. Baldoni, J. M. Marquez-Barja, P. Hellinckx, and S. Latré, "Requirements and Specifications for the Orchestration of Network Intelligence in 6G," Dec. 2021. doi:https://dx.doi.org/10.5281/zenodo.5767861.

[12] ETSI, "Experiential Networked Intelligence (ENI); Overview of Prominent Control Loop Architectures," *ETSI GR ENI 017 V2.1.1*, 2021. Online [Available]: https://www.etsi.org/deliver/etsi_gr/ENI/001_099/017/02.01.01_60/gr_ENI017v020101p.pdf, Date accessed: 2023-05-01.

[13] 3GPP, "Technical Specification Group Services and System Aspects; Procedures for the 5G System (5GS) Stage 2," *3GPP TS 23.502 V16.6.0*, 2020. Online [Available]: https://www.3gpp.org/ftp//Specs/archive/23_series/23.502/, Date accessed: 2023-05-01.

[14] 3GPP, "Application layer support for Vehicle-to-Everything (V2X) services; Functional architecture and information flows ," *3GPP Technical specification (TS)*, 2019. Online [Available]: https://www.3gpp.org/ftp/Specs/archive/23_series/23.286/, Date accessed: 2023-05-01.

[15] Z. Li, H. Li, and L. Meng, "Model Compression for Deep Neural Networks: A Survey," *Computers 2023*, vol. 60, no. 3, 2022. doi: https://doi.org/10.3390/computers12030060.