Time-based coordination in Intent-driven management for Vehicular Service Orchestration

Raúl Cuervo Bello, Nina Slamnik Kriještorac, Johann M. Marquez-Barja
University of Antwerp & imec, Internet & Data Lab (IDLab), Antwerp, Belgium
Email:{raul.cuervobello, nina.slamnik, johann.marquez-barja}@imec.be

***Abstract*—Intent-driven network management has become an important part of autonomous systems in Beyond 5G (B5G) towards Sixth-Generation (6G) networks, by enabling flexibility in the interaction among applications, operators and users. Intents play an important role in the communication of road users like autonomous vehicles and pedestrians to edge computing services. As sensor technologies for modern vehicles are cheaper, smaller, diverse and computing capable, more demand for applications and services on the road is increasing. A flexible intent interpretation and coordination are needed to deal with the dynamic environment and constantly changing goals. This paper presents a proof-of-concept of Zero-touch Network and Service Management (ZSM) for vehicular communication services, using an Intent Management Entity (IME) to translate user objectives into actionable directives. This paper describes a realistic testbed setup at the Smart Highway, where a Deep Reinforcement Learning (DRL) algorithm is used to optimize the selection of Roadside Units (RSUs) for service orchestration. This paper also discusses the challenges and opportunities of enhancing the IME with time-based intent coordination, using Artificial Intelligence and Machine Learning (AI/ML) techniques to estimate the waiting time and priority in intent coordination. The paper aims to demonstrate the benefits of ZSM and Intent-driven Management for vehicular edge computing and B5G/6G autonomous network management frameworks.**

***Index Terms*—zero-touch service management, intent-driven management, testbed, Smart Highway, orchestration, automotive services, vehicular communication**

I. Introduction

Advances in vehicular technology make it possible for vehicles to assist drivers thanks to the inclusion of sensor technologies like cameras, Light Detection and Ranging (LiDAR), radar and microprocessing capabilities such as On-board Units (OBUs), which allow autonomous driving or collision prevention [1]. When data from different sensors are integrated, this is called *sensor fusion*, and it enables greater accuracy in data processing such as object detection for collision avoidance [2]. Autonomous vehicles can make better and safer choices by combining different sensors to compensate for the gaps in those devices that might underperform in certain kinds of scenarios [3]. As the number of sensors in the car increases, the chances of making safer decisions increase in real-time driving circumstances that involve other vehicles [4], people or unpredicted obstacles like animals on the road. Sensor technologies are becoming more present in modern vehicles as they get less expensive and smaller [2]. The information collected from the surrounding environment by the sensors generates a large amount of data [1], including the data received from the nearby edge computing facilities, enabling the users to access a wider range of applications such as real-time features and analytics that can support critical services like collision avoidance as part of smart traffic management which have a strong impact on the safety of Vulnerable Road Users (VRUs). The increase of these functionalities in modern vehicles brings with it a greater demand for computational and communication resources and faster and more reliable connectivity.

To compensate for the workloads of vehicular technology applications, vehicular services can be deployed closer to the users from the cloud and edge computing entities providers [5], such as Roadside Units (RSUs) that can offload the computational burden of the user equipment and be supported by communication standards like Intelligent Transportation System (ITS-G5) and Vehicle-to-Everything (V2X). An RSU can also be interconnected to other RSUs and enable the network to enlarge the scope where vehicular services can operate. The downside of having these supporting units is the increment of the complexity of the network, as their services have to be properly managed optimally and efficiently to cope with the dynamicity of the demand [5]. Nevertheless, the available resources at the network edge, where the Multi-Access Edge Computing (MEC) units are located, may have limited and scattered resources in the diverse and complex B5G environment. To meet the real demands, the resources should be smartly distributed to where they are required. The coordination of this process among the network services and the vehicles can benefit from Zero-touch Network and Service Management (ZSM) techniques, which can be a collection of services that together offer functionalities for autonomous network and service management [6].

According to the European Telecommunications Standards Institute (ETSI) ZSM Group [7], intent-based interactions in autonomous network management frameworks are part of ZSM, and the goals and expected behavior are intent-defined as agnostic instructional language that enables services to interact with each other independently of the diverse technologies from the vendors, providers and operators. The management of intents includes the interpretation of the requests but also an optime coordination for its fulfillment to satisfy the users' Quality of Service (QoS). As the ZSM reference architecture [8] is service-based, intent-driven management has to perform cooperatively like an Intent Management Entity (IME) within the ZSM framework to handle the intent requests from the Management Service (MnS) producer who is the intent owner, and then generate the intent requests (MnS consumer) to a further level across the domains. IMEs can also contain knowledge-driven decision-making processes [7]. In practical terms, we designed a ZSM-based Network Orchestration Proof-of-Concept (PoC) for vehicular communication services, where decision-making processes optimize the consumption of net-

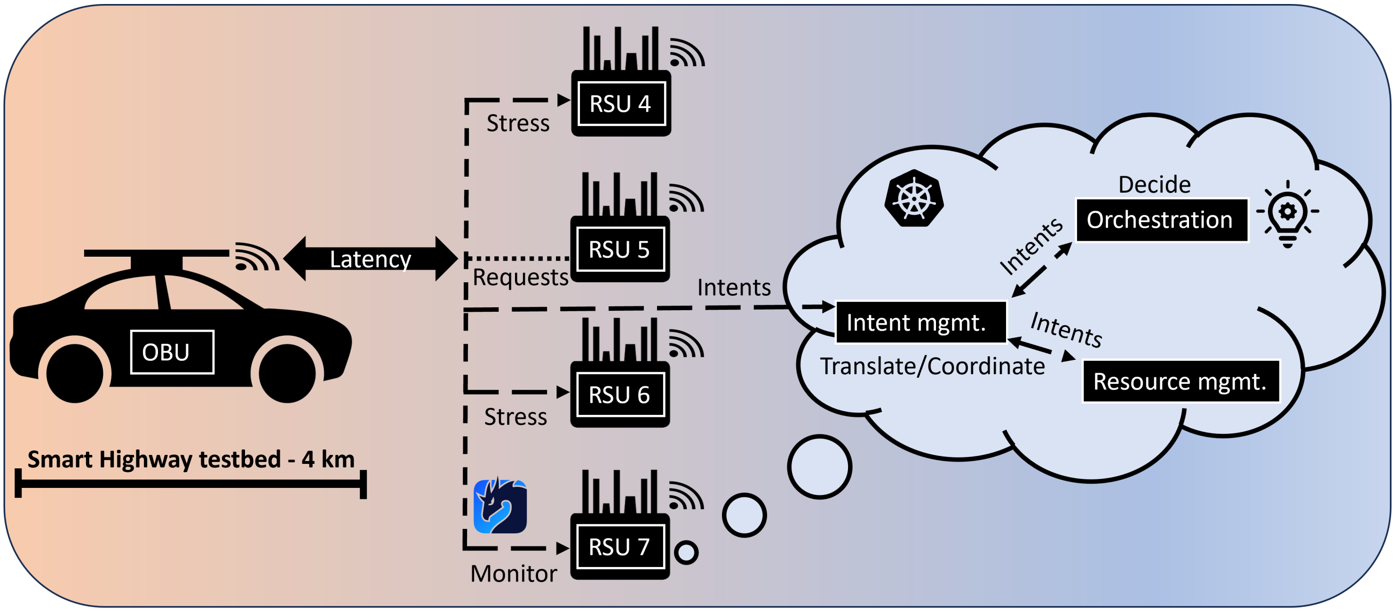

Fig. 1: ZSM-based Vehicular Network Orchestration PoC at the Smart Highway testbed.

work resources to manage and ensure demand stability. The PoC runs in a realistic environment spanning 4km of the Smart Highway testbed in Antwerp, Belgium, as presented in Figure 1. A basic intent interpretation is used in the current model, which motivates our goal to expand the capabilities of the current PoC to optimally coordinate the interpretation and execution of intents. To accomplish this, we propose an IME in charge of the interpretation that can be used by different intent owners like demanding users/operators or applications placed in a high-level domain and another for the coordination of the intents at the domain where the decision-making process takes place. Currently, a try/fail approach is used to ensure the fulfillment of the request in case of failure but with a pre-defined number of trials based on priority. We also aim to reduce the consumption of resources dedicated to coordinating the execution of the intents, by the use of time-based criteria to reduce the overflowing queues [9], where waiting intents can be sorted by the use of advanced prediction methods that are based on Artificial Intelligence and Machine Learning (AI/ML) techniques.

II. ZSM-based Network Orchestration PoC

The deployment of a PoC aims to test and validate Zero-touch services within the realistic conditions provided by the Smart Highway testbed, which stands as a Cooperative Intelligent Transport System (C-ITS). In this testbed, we make use of four out of seven RSUs that are available from a section of the E313 highway in Antwerp, Belgium. In every edge unit, wireless communications are provided by Cohda MK5 and MK6c over ITS-G5 and Long-Term Evolution (LTE) with a PC5 interface in the band 5.9GHz and a Uu interface at 3.5GHz and a Peplink 5G[1]. For computing processing, General Purpose Computing Units (GPCUs) are provided by Intel Xeon 8 Cores and 32 GB RAM. Energy management is possible through a Power Control Unit that is available in the RSU components. An additional mobile unit is deployed in a BMW X5 xDrive25d LO provided with similar equipment as the RSUs. The aforementioned devices that are contained in the computing units, supply facilities for sensing, hosting services and communication infrastructure to collect data in a highly realistic environment. The hosting services are containerized pods, powered by Kubernetes and implemented throughout the RSUs. These services can be accessed remotely during tests and experimentation for monitoring computing and network resources, and then the orchestration of vehicular services. To exchange the metrics data from the computing units we used Zenoh[2] as a publish/subscribe protocol that allows the services to retain an optimal level of time and space efficiency, thanks to its low latency and high throughput.

The Zero-touch services implemented in the Smart Highway testbed are enhanced by AI/ML-based functions and algorithms, enabling the orchestration service by improving the decision-making process. The intent-based interaction among monitoring, decision-making, and management services introduces flexibility, enabling a scalable and expandable range of solutions. Unlike relying on static pre-defined actions such as service scalation or termination, this approach enhances adaptability within the system. In Figure 1, our PoC setup consists of i) 3 RSUs to host vehicular network services, ii) 1 RSU for metrics collection and orchestration of the services, iii) 1 RSU for UE emulation to send requests and induce stress over the available services for testing purposes and iv) 1 The mobile OBU collects metrics from the rest of the RSUs to measure End-to-End (E2E) latency of the services, that is correlated to the consumption of computing resources like CPU and memory, which is an indicator of much workload a RSU is taking. All the RSUs have communication capabilities between them. The dynamic discovery features enabled by Zenoh enable the automatic subscription of the RSUs to the orchestration hub when the datasets of metrics are published. These datasets

[1]Peplink 5G: https://www.peplink.com/technology/what-is-5g-with-peplink/

[2]Zenoh: https://zenoh.io/

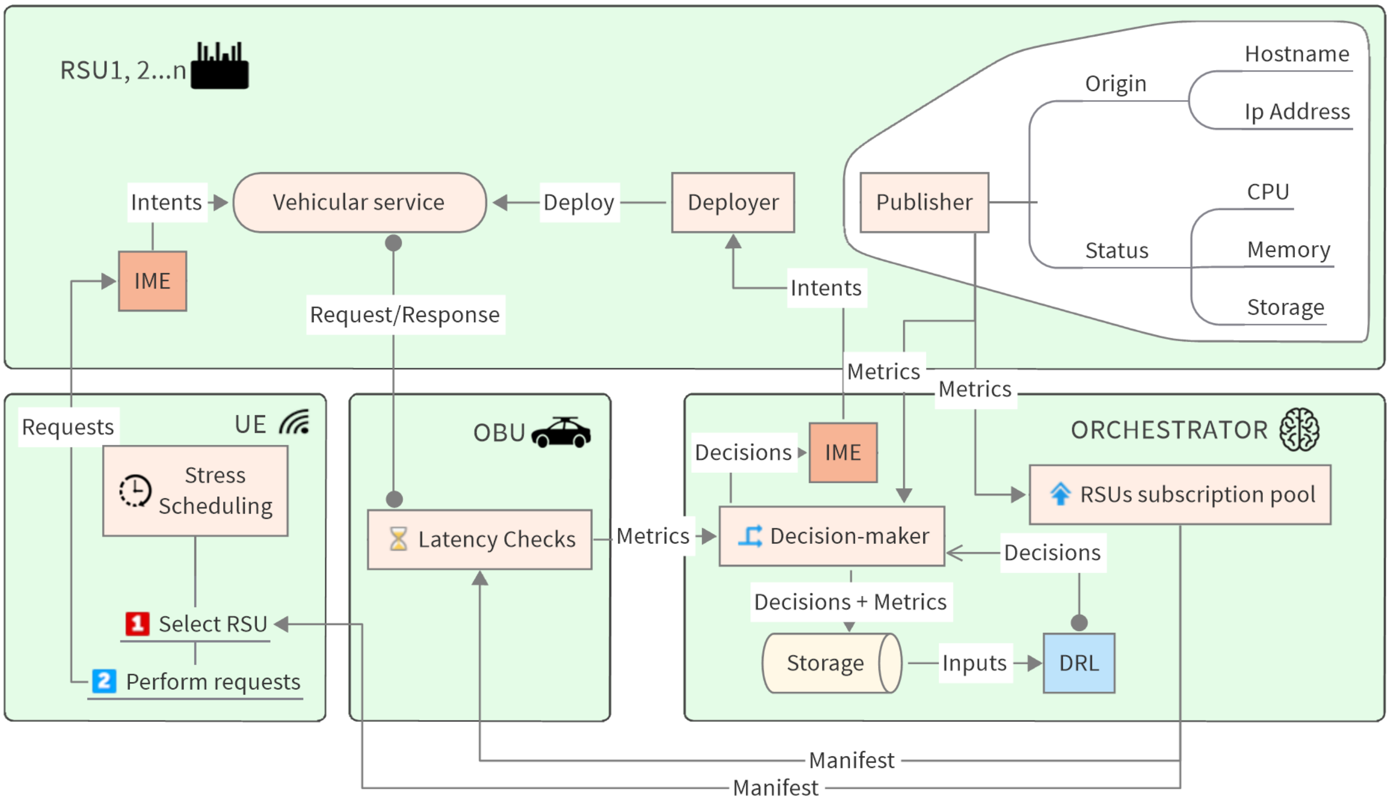

Fig. 2: ZSM-based Network Orchestration Flow.

include information about the origin entity like hostname and IP address. The origin data is sent and managed by the subscription pool which makes them available to the rest of the applications as shown in Figure 2. No manual configuration is needed before the execution of the PoC, providing flexibility to the setup and real-time feedback from the running services.

The execution of the PoC proceeds as follows displayed in Figure 2:

1) The publisher module of the hosting RSUs sends status values about performance status such as CPU load and Memory use. The dataset also hosts identification properties.
2) The RSUs subscription pool includes the newly arrived RSU identity and shares it with the rest of the applications involved in the PoC.
3) The OBU performs E2E latency checks among the RSUs.
4) The decision-making algorithm processes the datasets sent by the publisher module from the RSUs and the OBU, making choices on the most suitable RSU based on the availability of its resources. The resulting decisions are stored and used to train a Deep Reinforcement Learning (DRL)-based decision-making algorithm.
5) The stress-inducing scheduler application sends requests to a random RSU that is periodically selected from the manifest shared by the subscription tool. This task is performed to emulate the behavior of the dynamic changes in demand of vehicular network resources, which trigger actions in the orchestration operations for balancing and adjusting the running services.

The obtained output decisions in a previous rules-based decision-making algorithm were used for training a DRL-based algorithm, that later is integrated as a Zero-touch service and is further consumed by the orchestrator. The RSUs are part of the distributed edge computing within the testbed under a dynamic and complex environment, where DRL has become a viable choice for decision-making processes, demonstrating its effectiveness in learning optimal strategies within complex network scenarios and adapting to changes in recent years [10]. In a recent study that focused on our decision-making algorithms, the performance of the trained DRL algorithm has been recently validated through simulation and experimentation against a rules-based algorithm and it is in the works of integration with the current PoC. To tackle the issue of RSU selection a Deep Q-Network (DQN) algorithm was customized, which is a widely recognized DRL approach. This is significant for users as it directly impacts the availability and performance of vehicular services during periods of saturation caused by ineffective demand management, emphasizing the need for an efficient RSU selection. The decisions are then expressed declaratively through an intent, incorporating the identity of the RSU with superior resource availability. This is essential for optimizing both network and service performance. In our current setup, we aim to expand the current basic approach of expressing and interpreting the intents, which reflects the expected outcome instead of a detailed instructional set for the solution.

III. Intent management entity for interpretation and coordination

A. Intent Management Entity

An IME component is essential for the implementation of Zero-touch Service (ZTS) within the PoC framework. It plays a crucial role in translating high-level objectives into actionable directives for the network infrastructure. The intricate nature of ZTSs, designed for automated management of diverse network functions, demands mechanisms to interpret and execute user intentions. IMEs have a knowledge base that contains the intent ontology [7] and are also found to be queue-based as the requests are coordinated based on priority and availability of resources [2]. The lifetime of an intent can depend on its priority, which determines the number of tries for its fulfillment. These trial/error approaches tend to increase the consumption of resources as the number of attempts and timing are predefined before they are queued along with the rest of the requests and consequently are fixed. These approaches can benefit from a more accurate determination of the lifetime of the intent by extending the criteria for priority classification. Machine reasoning plays a key role in intent management, with its capability to understand abstract concepts from diverse domains and provide precise, specialized conclusions based on precedent as it can contribute to the quantification of risk and uncertainty, making decisions in circumstances of conflicting goals and new situations [7]. In this section, we address two crucial features of IME: intent interpretation and intent coordination. In the case of interpretation operations, they are mainly located in the first line of the ZSM architecture (See Figure 1) where the interaction of the users or operators is made through high-level instructions. Regarding intent coordination, it can be located near the inner domains of the ZSM architecture, managing the resulting outputs of the decision-making processes where more intents are generated with more specific instructions and targets.

B. Intent-based interpreter

Intents can serve various purposes, such as configuring or overseeing the performance of the network or service, and more. To cope with these purposes, an enhanced intent interpreter algorithm needs to be designed to provide an interface that receives and delivers instructions from high-level intents and compiles them into more detailed and technically elaborated intents that can be used by the destined orchestrators/controllers that will be applying changes and configurations to comply with requirements from the intent. Following the specifications from ETSI [7], the design of the intent interpreter should ensure that intents are expressed in terms of the intent owner, which does not specify how the desired service is to be realized. A pre-process can be employed to deconstruct natural language intents where the intents are formatted into the structure that is expected by the IME. These intents are later submitted in the network as standardized information objects. Among related works that propose different approaches for the development of intent interpreters, a more generic proposal comes from McNamara et al., as they present a flexible interpreter for intent realization [11]. Firstly, they identify a mathematical and a model-based as a means of representation for intents. Secondly, a flexible interpreter takes care of the aforementioned representations through the following stages: functionality templates, intent matching and translation.

C. Intent coordination

Beyond 5G (B5G) and Sixth-Generation (6G) future networks are essentially intent-driven and proper coordination of the realization of the intents contributes to their success and eventually the performance of autonomous network management frameworks. In the search for practices in intent coordination, we have found increasing attention to the subject from researchers and network operators as well. In this subsection, we present and discuss different approaches used to manage the execution of intents.

In related work, Perepu et al. [12] propose a method based on Multi-agent Reinforcement Learning (MARL) for coordination and conflict management among loops to achieve intent-based management. Following this approach, they aim to tackle the challenges of the lack of compute optimization at runtime in dynamic environments where the goals change continuously. By solving the conflicting intents, the MARL agents can enable cooperation in the loops by prioritizing the important Key Performance Indicators (KPIs) and reducing human intervention. The paper concludes with the validation of the proposed MARL model through experimentation, where the results showed that MARL-based agents can learn the *Plan to Coordinate* to achieve an optimal global trade-off during the training phase, enabling them to effectively handle conflicts and encourage cooperation to maximize the achievement of the global goal. TianZhang He et al [13], propose an approach to orchestrate networking and computing resources based on user requirements while filling the gaps in existing Vehicular Edge Computing (VEC) applications. This involves developing necessary algorithms that take into account both the computing and networking needs of the applications. As a result of their work, two algorithms for intent resolution were designed.

Priority-Aware Intent Installation: This algorithm periodically checks the intents associated with a suspended event. If an intent cannot be satisfied, the algorithm switches the state of the event to 'Failed' and schedules it for reinstallation, with a threshold of 3 attempts.

Location-Aware Mapping: Consists of a mapping microservice dedicated to determining the best path for allocating virtual nodes by taking location contains from the intents. The algorithm later sorts the searched nodes based on the depth and the distance, which is particularly useful for mobile end-users like vehicles and VRUs where the location of the virtual node is expected to change in time. This would minimize the impact on node reallocation in dynamic environments that imply mobility-related scenarios.

In the case of the Priority-Aware Intent Installation algorithm, the retry threshold can be dynamically adjusted based on current edge conditions. In dynamic environments, a low threshold may lead to a high failure ratio, while a high threshold can increase intent reinstallation and processing time. Additionally, a larger retry threshold may enhance the acceptance rate for high-priority intents. Retry attempts are traditionally set at a fixed number, such as three tries. However, there is an opportunity for more intelligent determination of retry parameters, including timeouts, by incorporating AI/ML support. This would allow for dynamic adjustments based on rules rather than relying solely on fixed numerical values, although is not exempt from challenges like extra computing efforts for prediction and mapping.

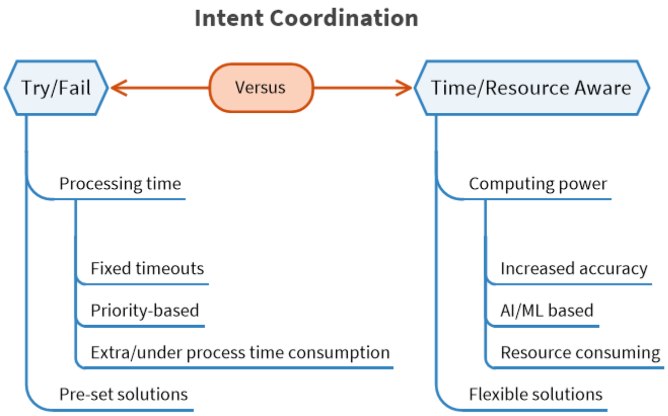

Fig. 3: Intent-driven threshold coordination approaches.

D. *Time-based coordination in intent-driven management*

If intent processing is not optimized, then the expected effect over performance would be contradictory. Instead of trying for re-compilation, the intent manager stands by for notification or for a queue of executions for when the conditions are suitable. The monitoring/update application can collect the requests and give them a timeout that can be pre-set or determined based on AI/ML-based time estimation. In Figure 3 we compare two different approaches regarding execution threshold in intent-coordination. AI/ML techniques can be used to prioritize orders in a queue by implementing a priority queue system. The prediction tool that can be used to estimate when a particular service is going to be executed based on a queue of pending actions and the current status of available resources is a queue management system. By incorporating an order queuing tool that specifically estimates waiting times into intent coordination, we can achieve more accurate sorting of pending intents. This enables the determination of estimated response times, ultimately reducing costs associated with holding unnecessary resources. The goal is to identify a suitable prediction tool within AI/ML that can estimate when a specific service will be executed based on a queue of pending actions and the real-time status of available resources. This approach aims to optimize service execution by providing accurate estimates based on dynamic factors, thereby improving overall efficiency in the queuing system.

IV. Conclusion and Future Work

In this paper, we introduce the performance of IMEs within the context of a ZSM and the crucial role these entities play in translating user objectives into actionable directives for optimal network management. We also described a PoC of a ZSM-based framework at the Smart Highway testbed that showcased the practical implementation of a ZSM framework, incorporating a DRL-based decision-making algorithm for vehicular network service orchestration. We also discussed the goals of expanding the PoC capabilities by the integration of IMEs enhanced with time-based intent queuing coordination. As part of future work, a systematic assessment of the proposed time-based intent coordination, in conjunction with the previously discussed coordination approaches, will be undertaken. This evaluation aims to provide valuable insights into the effectiveness and performance of time-based intent coordination, offering a nuanced understanding of its impact on the overall orchestration and management of vehicular network services within the evolving landscape of advanced network management frameworks.

Acknowledgement

This work has been performed within the European projects: Horizon 2020 Vital-5G project (Grand Agreement No. 101016567) and SNS project: TrialsNet (Grant Agreement No. 101095871).

References

[1] F. A. Butt, J. N. Chattha, J. Ahmad, M. U. Zia, M. Rizwan, and I. H. Naqvi, "On the integration of enabling wireless technologies and sensor fusion for next-generation connected and autonomous vehicles," *IEEE Access*, vol. 10, pp. 14643–14668, 2022. doi: https://doi.org/10.1109/ACCESS.2022.3145972.

[2] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," *Sensors*, vol. 21, no. 6, 2021. doi: https://doi.org/10.3390/s21062140.

[3] T. Nguyen and M. Yoo, "Fusing lidar sensor and rgb camera for object detection in autonomous vehicle with fuzzy logic approach," in *2021 International Conference on Information Networking (ICOIN)*, pp. 788–791, 2021. doi: https://doi.org/10.1109/ICOIN50884.2021.9334015.

[4] N. Slamnik-Kriještorac, M. Camelo, C.-Y. Chang, P. Soto, L. Cominardi, D. De Vleeschauwer, S. Latré, and J. M. Marquez-Barja, "Ai-empowered management and orchestration of vehicular systems in the beyond 5g era," *IEEE Network*, vol. 37, no. 4, pp. 305–313, 2023. doi: https://doi.org/10.1109/MNET.008.2300024.

[5] A. Singh, G. S. Aujla, and R. S. Bali, "Intent-based network for data dissemination in software-defined vehicular edge computing," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 5310–5318, 2021. doi: https://doi.org/10.1109/TITS.2020.3002349.

[6] M. Liyanage, Q.-V. Pham, K. Dev, S. Bhattacharya, P. K. R. Maddikunta, T. R. Gadekallu, and G. Yenduri, "A survey on zero touch network and service management (zsm) for 5g and beyond networks," *Journal of Network and Computer Applications*, vol. 203, 2022. doi: https://doi.org/10.1016/j.jnca.2022.103362.

[7] E. T. S. Institute, *Zero-touch network and Service Management (ZSM); Intent-driven autonomous networks; Generic aspects*. 2023. Online [Available]: https://www.etsi.org/deliver/etsi_gr/ZSM/001_099/011/01.01.01_60/gr_ZSM011v010101p.pdf, Date accessed: 2023-12-14.

[8] ETSI Industry Specification Group (ISG), "Zero-touch network and Service Management (ZSM); Reference Architecture." ETSI GS ZSM 002 V1.1.1 (2019-08), 2019. Online [Available]: https://www.etsi.org/deliver/etsi_gs/ZSM/001_099/002/01.01.01_60/gs_ZSM002v010101p.pdf, Date accessed: 2023-09-07.

[9] D. Terekhov, T. T. Tran, D. G. Down, and J. C. Beck, "Integrating queueing theory and scheduling for dynamic scheduling problems," *J. Artif. Intell. Res.*, vol. 50, pp. 535–572, 2014. doi: https://doi.org/10.1613/jair.4278.

[10] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3133–3174, 2019. doi: https://doi.org/10.1109/COMST.2019.2916583.

[11] J. McNamara, E. Aumayr, L. Fallon, and E. Fallon, "A flexible interpreter for intent realisation," in *NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium*, pp. 1–6, 2022. doi: https://doi.org/10.1109/NOMS54207.2022.9789910.

[12] S. K. Perepu, J. P. Martins, R. S. S, and K. Dey, "Intent-based multi-agent reinforcement learning for service assurance in cellular networks," in *GLOBECOM 2022 - 2022 IEEE Global Communications Conference*, pp. 2879–2884, 2022. doi: https://doi.org/10.1109/GLOBECOM48099.2022.10001426.

[13] T. He, A. N. Toosi, N. Akbari, M. T. Islam, and M. A. Cheema, "An intent-based framework for vehicular edge computing," in *2023 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pp. 121–130, 2023. doi: https://doi.org/10.1109/PERCOM56429.2023.10099081.