Received 12 July 2023, accepted 31 July 2023, date of publication 2 August 2023, date of current version 9 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3301330

RESEARCH ARTICLE

5G on the Roads: Latency-Optimized Federated Analytics in the Vehicular Edge

LÁSZLÓ TOKA[1,2,3], (Member, IEEE), MÁRK KONRÁD[1], (Graduate Student Member, IEEE), ISTVÁN PELLE[1,2,3,4], BALÁZS SONKOLY[1,2,3], MARCELL SZABÓ[1], BHAVISHYA SHARMA[5], SHASHWAT KUMAR[5], MADHURI ANNAVAZZALA[5], SREE TEJA DEEKSHITULA[5], AND A. ANTONY FRANKLIN[5], (Senior Member, IEEE)

[1]HSN Laboratory, Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, 1111 Budapest, Hungary
[2]ELKH-BME Cloud Applications Research Group, 1111 Budapest, Hungary
[3]MTA-BME Network Softwarization Research Group, 1111 Budapest, Hungary
[4]EIT Digital Doctoral School, 1111 Budapest, Hungary
[5]Indian Institute of Technology Hyderabad, Hyderabad 502284, India

Corresponding author: László Toka (toka.laszlo@vik.bme.hu)

This work was supported in part by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund under the FK_20 Funding Scheme under Project 135074; in part by the 2019-2.1.13-TÉT_IN Funding Scheme under Grant 2019-2.1.13-TÉT_IN-2020-00021; in part by the 2019-2.1.11-TÉT Funding Scheme under Grant 2019-2.1.11-TÉT-2020-00183; in part by the Strengthening the European Institute of Innovation and Technology (EIT) Digital Knowledge Innovation Community in Hungary under Grant 2021-1.2.1-EIT-KIC-2021-00006; in part by the New National Excellence Program of the Ministry for Culture and Innovation from the Source of the National Research, Development and Innovation Fund under Grant ÚNKP-22-5-BME-317; and in part by the Department of Science and Technology (International Cooperation Division), Government of India, through the project "Autonomous Driving Enabling fog Computing Platform With Edge Cloud Orchestration and Edge Analytics." The work of László Toka was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

ABSTRACT Coordination among vehicular actors becomes increasingly important at the dawn of autonomous driving. With communication serving as the basis for this process, latency emerges as a critical limiting factor in information gathering, processing, and redistribution. While these processes have further implications on data privacy, they are also fundamental in safety and efficiency aspects. In this work, we target exactly these areas: we propose a privacy-preserving system for collecting and sharing data in high-mobility automotive environments that aims to minimize the latency of these processes. Namely, we focus on keeping high definition maps (highly accurate environmental and road maps with dynamic information) up-to-date in a crowd-sourced fashion. We employ federated analytics for privacy-preserving, low-latency, scalable processing and data distribution running over a two-tiered infrastructural layout consisting of mobile vehicular nodes and static nodes leveraging the low latency, high throughput and broadcast capabilities of the 5G edge. We take advantage of this setup by proposing queuing theory based analytical models and optimizations to minimize information delivery latency. As our numerical simulations over wide parameter-ranges indicate, the latency of timely data distribution can be decreased only with careful system planning and 5G infrastructure. We obtain the optimal latency characteristics in densely populated central metropolitan scenarios when Gb/s uplink speeds are achievable and the coverage area (map segment size) can reach a diameter of 1 km.

INDEX TERMS Edge platform, end-to-end latency, federated analysis, vehicular communication.

I. INTRODUCTION

Based on current automotive trends and developments, it is evident that autonomous vehicles and advanced driver assistance systems (ADAS) are transforming the industry. Alongside the increase of onboard computing power [1], [2] that creates a basis for such advanced features, we can also observe a gradual shift from running vehicular control tasks solely using onboard electronics to forming a cooperation with remote computing infrastructure [1], [2] and

The associate editor coordinating the review of this manuscript and approving it for publication was Angelo Trotta.

roadside units (RSU) [3] via vehicle-to-infrastructure (V2I) communication. These extend both vehicular sensing and computing powers. In our view, this evolution will lead to platforms and applications that provide built-in support for efficiently handling the infrastructure hierarchy as a continuum ranging from onboard devices through edge/RSU units up to the cloud. It is also expected that these factors will lead to such potentially high-impact benefits as increased safety, improved traffic flow, and reduced environmental impact. However, before we can reap these benefits, there are still a plethora of hurdles to be solved on the technical, regulatory, and societal levels. On the technological side, the compute-intensiveness and latency-sensitivity of control tasks and the information sharing between participants are present challenges. This latter aspect is also of concern from a regulatory or societal point of view as data security and privacy implications become more and more important with the growing amount and kind of data that are being collected by the ever-increasing number of sensors located on our mobile vehicles and in the stationary units placed alongside the road infrastructure.

In recent years, we could see two major areas that helped improve latency: reducing transmission delay (by simultaneously providing higher throughput for more users) and bringing compute resources closer to end-users. 5G cellular networks leverage both. They enable high-speed data transfers with low transmission latency even for a higher number of mobile users to be served. They also provide options for running computations close to the edge of the network. In a more general sense, similarly to cloud computing, edge computing offers a model for delivering computing resources to end-users, but contrary to the cloud, potentially much closer to where data is being generated [4], [5], [6]. Leveraging these computing resources provides faster processing of data (compared to on-device processing) and reduces network latency. These aspects are especially important for applications that require real-time processing, such as those appearing in automotive cases. While multiple players in the automotive industry state that safety-critical computations should be performed on board of the vehicles in the near future, they clearly identify (beyond) 5G networks as crucial enablers for realizing self-driving features. The 5G Automotive Association (5GAA) prescribes a service level latency of 100 ms for the *Real-Time Situational Awareness and High-Definition Maps: Hazardous Location Warning* use-case [7] which is within current 5G capabilities.

While edge computing provides the resources for running tasks close to end-users, data collection, delivery, and processing methods can also highly improve the latency of services provided to the users. Federated analytics (FA) [8] is such a novel distributed, privacy-preserving approach that combines data from multiple sources to create a comprehensive view. It enables data analysis and gaining of insights from distributed data without having to move the respective data from its original location. In federated analytics, data is analyzed locally on devices where it is generated and stored, and solely the results are shared and aggregated on a central server to perform the actual analysis. Due to this behavior, it can increase data processing efficiency on large and distributed data sets as a lower amount of data needs to be transferred and stored in a central location. By retaining the control of potentially sensitive data with its original owner, federated analytics also limits the number of entities having direct access to the data, thus increasing security and privacy aspects.

In our view, the ensemble of these technologies gives us the right toolkit for providing advanced vehicular support functions. We argue (and later demonstrate) that in such scenarios, the combination of 5G and federated analytics can serve as a good basis for high-mobility automotive environments that connect static and moving units. As we see, the concept maps well to a two-tiered infrastructure where vehicles collect data, while 5G edge units aggregate and distribute the insights from localized data without breaching data privacy and without the need for vehicle-to-vehicle (V2V) communication. Using federated analytics, we assume the exchange of only insights instead of uploading raw data which reduces transmitted data sizes and consecutively total service latency as well. In order to harness the concept to its full potential, we need to take into account data locality, the diverse compute capabilities and latency characteristics of each hierarchy level. Data collection and processing have to be performed at the most adequate level to fully exploit the multitude of vehicular sensors and increased processing power. These can enable constant frequent updates on the behavior of traffic participants and road conditions which can pave the way for assisted or autonomous driving scenarios [2] preparing for AI applications.

As we see, an application that is a natural fit for such a platform is the maintenance of multilayered high definition (HD) maps. Such HD maps provide detailed, up-to-date information about the surrounding environment, where data changes at a different dynamicity rate in each layer of the map. HD maps allow for advanced safety and navigation capabilities and are especially beneficial for vehicles in autonomous or semi-autonomous driving modes. Leveraging edge computing and federated analytics can improve the accuracy of (non-mission-critical) map data via reduced latency, data collection from various sources, and distributed processing of this data.

Consequently, our contributions are the following. *i)* We introduce a federated analytics-based solution for collecting and aggregating data originating from vehicles. *ii)* We propose an analytical model that expresses the end-to-end latency of channeling information through the system. Our model that leverages queuing theory is used for providing a closed formula to determine the response time of our system. *iii)* We extend this with our own information value metric that provides diminishing returns as the number of clients in a 5G cell increases. It takes into account the number of clients in a 5G cell and the importance of the data collected from each client. *iv)* We use our model to analyze the performance of the system and optimize its parameters to improve the efficiency of HD map data collection and processing. *v)* We also provide a multi-aspect numerical evaluation of our solution in low-latency information delivery to help system dimensioning using a simulation built on top of a well-known mobility framework.

To discuss these, we structure our paper in the following way. We review related work in Section II, discussing vehicular information delivery. Then we elaborate on the envisioned system and HD maps service in Section III. Later we detail our system model using the federated analytics-based upstream scheme in Section IV. We highlight important aspects of our evaluation environment, scenarios, and results in Section V. Finally, in Section VI, we draw conclusions.

II. RELATED WORK

This work is an expansion of our previous 4-page-long conference short paper [9] which we heavily reworked providing deeper insights into our motivations and HD map management use-case. We have also provided a deeper explanation of our theoretical model, optimization, and numeric evaluations which we extended with new results as well.

In the rest of this section, we briefly look over different areas: We analyze various recent solutions regarding HD map building and federated methods used in automotive contexts. As our system applies a collaborative map creation process by offloading computation tasks to edge resources, we review related work from this respective area as well. Finally, we visit the privacy aspects of the use of federated methods in vehicular settings.

Existing works on content distribution in vehicular networks only consider static content and fail to address the dynamic information exchange present in cases involving HD maps. The HD map distribution problem differs from normal content distribution due to frequent data changes that must be disseminated to all caching locations and end clients. Therefore distributing the HD map generates periodic traffic not only in access networks, but on the backhaul connecting to the RSUs or edge nodes too.

The main focus of Xie et al. [10] is reducing the power consumption during the distribution of high definition maps. They propose an algorithm that ensures an RSU only serves a vehicle if the energy required for basic movement and receiving data from the RSU is less than the remaining energy of the vehicle. To provide the HD map service, the system distributes the data proportionally to the available vehicle power. Wu et al. [11] suggest a policy that jointly controls power and assigns spectrum to maximize the data rate of the overall network for HD map distribution. They analyze the effect of the interference present in the system on data transmission and create a model to describe the problem of controlling this interference in HD map distribution over vehicle-to-everything (V2X) communications. They recommend the cooperative transfer of HD maps through V2I and V2V delivery by splitting the HD map into several blocks depending on the infrastructural environment and the volume of the data to be handled. Xu et al. [12] present a study on HD map caching for autonomous driving in vehicular networks, where vehicle requests and trajectories are unknown. The algorithm that they propose defines a reward function built on the history of requesting tile data recorded by the system.

Peng et al. suggest a new architecture that combines the Multi-access Edge Computing (MEC) and Software Defined Networking (SDN) concepts to facilitate autonomous driving assisted by HD maps [13]. In order to achieve efficient scalability and utilization of network and compute resources, the proposed two-tier server structure consists of cloud and MEC servers where services and applications are deployed on the MEC server at the network edge leveraging Network Function Virtualization (NFV) techniques. Zhang et al. also offer a framework for HD map applications based on MEC incorporating the application and functional modes and modules, the workflow of distributing HD map data, as well as the communication process between MEC and the autonomous vehicular client entities [14].

Federated analytics (or analysis) is a young concept introduced only a couple of years ago. We see a similar pattern in its application to its oldest sibling, federated learning: as time passes, it finds its way to more and more areas of application. While it has already been used in many use-cases since its introduction, analyzing vehicular data has not yet become prevalent and surveys regarding this field are also missing and, to the best of our knowledge, it has not yet been leveraged conjoined with 5G networking and edge computing in a vehicular setting. Exploratory works are present applying federated learning schemes in diverse V2V or V2I settings [15]. Within these, prominent areas are monitoring driver activity [16], [17], predicting steering wheel angle [18], [19], vehicle trajectory (sometimes with collision avoidance) [20], [21], [22], [23] and object detection [24], [25], [26], [27], [28], [29], [30]. Certain studies employ (beyond) 5G technologies [31] while others investigate performance in real or in simulated environments [32] clearly identifying the viability of the combination of these technologies, but also citing multiple open questions [33]. Liu et al. [34] offer a hierarchical federated learning system where they define and evaluate an algorithm that is capable of performing partial federated aggregations with vehicular clients connecting to cloud servers via an intermediate edge layer. They combine edge and cloud capabilities in order to find a balance between communication- (with edge nodes) and computation-efficiency (in the cloud) and conclude that model training time and the client energy consumption can be reduced compared to an architecture without an edge layer. A further study [35] complements this by investigating the effects of multiple intermediate layers on an HD map building and distribution use-case. The authors find that increasing the number of layers can accommodate more clients without significantly increasing serving latency.

Offloading computation tasks from resource-limited end devices to cloud, edge or fog computing nodes has become a well-studied area in recent years [36]. While multiple works relate to different aspects of our work, in our view, they do not cover the composition that we target. Here, we briefly review a few of these. In a hybrid telecommunication core-network-to-edge environment Moubayed et al. [37] describe the placement of V2X services as a binary integer linear programming problem minimizing the average end-to-end delay of service instances. Another study [38] gives a practical implementation for the relocation of the components of a vehicle remote control application with low latency overhead using an edge computing environment with possibilities for leveraging cloud infrastructure as well. Neither of these incorporate AI components, although, the second one uses an application performing object detection as a use-case. Wang et al. [39] model the vehicle-to-vehicle offloading process of application components as a Markov decision process minimizing the average completion latency of the application without utilizing the edge infrastructure or learning methods. A further work [40] leverages a learning-based approach for performing a similar task. Chen et al. [41] focus on energy consumption aspects in a prospective 6G environment targeting application component offloading to MEC servers via stochastic optimization.

As vehicles can generate huge amounts of data to be shared, data privacy concerns might arise as well. Regulatory bodies follow the technological advancement and fine-tune the rules that govern what data can be shared and how. To identify two major players, in the US, the National Highway Traffic Safety Administration (NHTSA) and Federal Trade Commission (FTC) are responsible for regulating the sector, while in the EU, the European Data Protection Board is responsible for making recommendations about data privacy even regarding vehicular data. They both agree that data sharing is to be adequately regulated when it can be used to reveal information tied to a person (i.e., driving style, covered distance, frequently traveled routes, precise addresses visited, [camera]sensor recordings). In the EU, such regulations can fall under the umbrella of the General Data Protection Regulation (GDPR). While federated learning or analysis, through sharing only insights with central (MEC) servers, hold the possibility of privacy-preserving data sharing, their traditional implementation might still suffer from training data mining when model privacy is not properly handled [22], attacks against servers or dealing with malicious servers or clients [42]. Leveraging these, attackers might be able to capture sensitive data or negatively influence the federated merge process. While protecting the privacy of the data generated by vehicular clients is not in the main focus of our work (it is merely an additional benefit of using federated analytics), we note, that there are multiple methods focusing on vehicular cases [22], [42], [43] or from other fields as well [44], [45] with which such issues can be addressed.

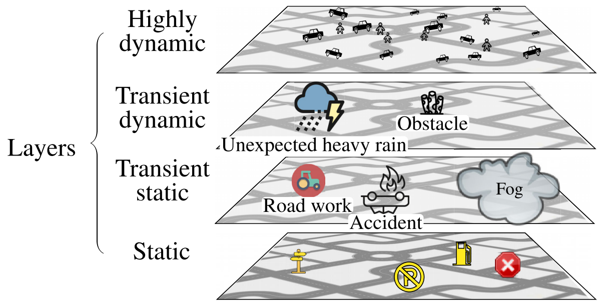

FIGURE 1. Layered HD map with static and dynamic information.

III. COLLECTING AND DISTRIBUTING HD MAP DATA

Autonomous driving is getting ever closer to availability to the public. However, real-time navigation tasks, especially in urban scenarios, are challenging, due to the limited range of sensors and inherited inaccuracies. This induces a growing demand for vehicles to have readily available, finely detailed information about their surroundings. High definition maps can be of great service in this regard [46] as they combine digital mapping technology and sensor data to achieve high levels of detail, accuracy, and update frequency when providing a comprehensive 3D view of a location. HD maps can contain precise and detailed information about the road network, including, for example, road geometry, lane markings, traffic signs, and lights together with other, more dynamically changing features. These maps are crucial for going beyond simple navigation assistance tasks and support full-fledged autonomous vehicles (AV), as they enable real-time vehicular navigation and decision-making by supplying the necessary up-to-date high-fidelity information with centimeter-level precision to these tasks. They also allow for advanced safety and real-time location-based services enabling abilities to perceive beyond sensor visibility, apply context awareness, and plan proactively.

The creation of an HD map typically involves three main phases. During the *i)data collection* phase, information originating from multiple sensors is gathered from a fleet of ordinary or specialized mapping vehicles using a combination of technologies. While (differential) GPS equipment provides accurate location and positioning information, high-resolution laser (LiDAR, Light Detection and Ranging) scans, and video cameras capture accurate 3D data of the road network, the surrounding environment and other relevant features, such as obstacles and traffic rules (e.g., signs and lane markings). In the *ii) processing phase*, the collected data is merged and processed by AI and machine learning algorithms to form a highly detailed and accurate map. As shown in Figure 1, such processing can aggregate multiple layers all with highly varying degrees of dynamicity to offer higher map accuracy enabling the making of safer decisions. A highly dynamic layer can supply real-time data on the surrounding environment, such as vulnerable road users or traffic congestion. This layer can encode data that changes on a second or minute timescale. A second, transient dynamic layer may provide less frequently changing information, for example, weather conditions that might change on a timescale of hours. A further transient static layer can record information that changes even less frequently, requiring weeks or months to be changed, e.g., road work and speed limits. The bottom-most static layer might contain the road map or layout, i.e., information that changes over the arc of multiple years and thus can be considered fixed. Finally, during the *iii) map delivery phase*, the map is distributed to the vehicular end-users.

Overall, HD maps are critical enablers of AVs, however, their creation is a complex and resource-intensive task, exacerbated by the high mobility of vehicles making the integration of several advanced technologies and specialized expertise quintessential. Maps are highly location specific, require the transmission of large data volumes, and have to convey changes to vehicles with low latency for efficiency and safety reasons. Vehicles in the same geographic location request the same map data (also known as a tile) which leads to the repetitive transmission of data volumes over the core network that stresses the capacity-limited backhaul links. Having multiple map layers allows for being more resource-efficient. Reducing the amount of transmitted and processed data by applying a layered structure rather than updating the entire map each time a change occurs, can lead to lower latency, reduce network load and make the map more scalable overall. The complexity of managing multiple layers, on the other hand, can require more advanced hardware and software capabilities to process and store the data. Fully onboard processing, traditional V2I solutions, data collection, and processing techniques are not designed to handle such scenarios, however.

We argue that these issues can be circumvented by building HD maps collaboratively with the process shown in Figure 2. Here vehicular clients first observe their surroundings (see step 1) to collect environmental data. Then they preprocess it locally (by performing object detection, and recognition [24], [25], [26], [27], [28], [29], [30], as well as trajectory tracking and prediction [20], [21], [22] – step 2) and upload the resulting insights over low-latency 5G wireless links to multi-access edge computing (MEC) servers at the vehicular network edge contributing to HD map layers matching the detected objects and client capabilities (step 3). These edge nodes then merge the insights of multiple vehicular clients periodically (step 4). For these steps, federated analytics can be an adequate solution. FA can leverage data locality, privacy, increased accuracy, and scalability. The set periodicity of the federated merges of insights can help make the development of on-vehicle applications easier. It can serve as a basis for creating standardized interfaces for accessing the federated merges and receiving updated map tiles preventing platform or vendor lock-in. The 5G edge can appear as a suitable infrastructure, providing the necessary computing capabilities for FA in the vicinity of the users (avoiding the traversal of the core network) and broadcast communication (as of Release 16 [47]) for redistributing the merged HD map tiles and layers (see step 5).

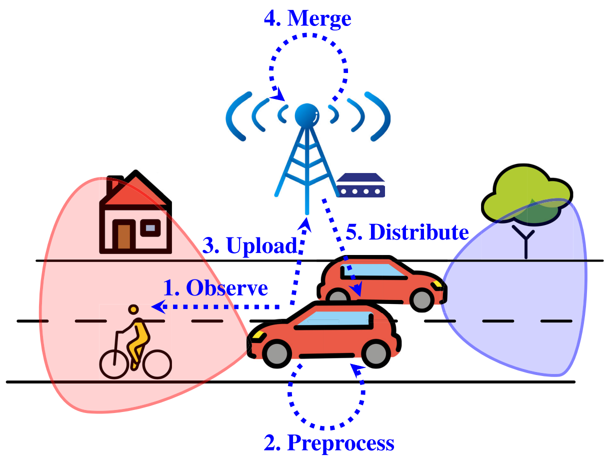

FIGURE 2. Schematics of system operation.

In this work, we make the case of deterministic HD map update periods in a combined edge computing and telco environment, and evaluate the toll that short update periods take on the system. However, other crowd-based information gathering use-cases can also benefit from the ideas presented here that utilize edge processing and data distribution over low-latency, high-bandwidth communication channels. Multi-user, collaborative extended reality (XR; encompassing holography, augmented, virtual and mixed reality), tactile internet and telemedicine applications can also be such targets, as they require environment detection and sharing of data among clients with strict low latency requirements [48], [49], [50], [51], [52], [53], [54], [55], [56].

IV. SYSTEM MODEL

Our proposed mathematical model that is able to describe our federated HD map distribution network is based on D/M/1 queuing systems. We discuss the model's formulas and necessary constraints along with a method to measure the value of information transmitted between clients and 5G edge locations in the following. We also leverage known numeric optimization algorithms in order to formulate guidelines and best practices for optimizing the performance of our system.

A. ON THE MOTIVATION FOR USING D/M/1 QUEUES

When closely inspecting the properties of our conceptual federated HD map distribution network, we make two major observations. First, at fixed, deterministic points in time, we merge data provided by the vehicular clients (step 4 in Figure 2). Second, the process according to which these clients provide data is affected by such factors as local construction, preprocessing and upload latency of FA insights to the serving edge node (steps 1–3 in Figure 2). These factors are determined by, e.g., locally available compute resources, network bandwidth and the properties of the radio channel. Collapsing these factors into a single stochastic process can simplify our system model, while it can also keep the necessary level of fidelity regarding our original assumptions.

Based on the above observations, we can conclude that neither M/M/1 nor M/D/1 queues are suitable for depicting the importance of fixed periodical trigger times of updating each HD map layer's data and stochastic insight generation. While the M/G/1 queue type might be an adequate option, it does not allow to formalize our problem in closed formulae, making the problem intractable. Thus we opted for using D/M/1 queues as a basis. In this work, we model the first factor as deterministic arrivals (denoted as β in the literature and as the T_t trigger time henceforth to better convey the underlying concept). For modeling the second factor, we use the following combined parameter:

$$T_w = T_p^j + T_r^j \; \forall j \simeq \max_j(T_p^j). \tag{1}$$

Here, the T_p^j *preparation time* incorporates all processes performed at the vehicular client j: environmental data collection, preprocessing, packaging, and uploading to the serving edge node. The T_r^j *response time* gives the delay for vehicle j to receive an update (merged, fully processed map data) from the serving edge node. This part is dependent on the T_p^j of all clients j collaborating on the same layer of a specific HD map tile and the delay of the merge operation. As the latter is considered to be fast, T_w can be approximated by the longest T_p^j: T_w is the *waiting time* which signifies the end-to-end latency in a traditional FA setup (for an illustration see Fig. 6.) We model this factor via D/M/1 queue's Markovian service time that is of exponential distribution. This is in line with networking-based queuing systems and, in our view, is able to adequately capture the nature of the combined local processing and upload processes of our use-case.

Based on the properties of multi-layered HD maps discussed in Section III, we assume that due to the highly different dynamicity of the information encoded in different map layers, different layer update trigger periods are also implied. While on-vehicle detection of the environment happens in real time, it makes sense if we combine these observations to, e.g., a highly dynamic and to a static HD map layer using different periodicities. We handle this by introducing independent closed D/M/1 queues for each map layer in our model. As noted in Section III, we assume that vehicular clients can contribute to multiple layers based on the objects that they identify and their capabilities. (In our opinion, the assignment of clients to layers is out of the scope of this work, thus we identify it to be explored in the future.) The independent queues account for both the uplink and downlink communication in their corresponding HD map tile.

B. OUR THEORETICAL MODEL FOR CALCULATING WAITING TIMES AND INFORMATION VALUES

In order to determine the value of information transmitted throughout the FA network incorporating vehicles and edge nodes, we first have to calculate the T_w waiting time of an HD map layer i in our D/M/1 queues. A prior study [57] gives a thorough analysis on D/M/1 queues with deterministic

TABLE 1. Mathematical notations used in our system model.

Notation	Description
N_c	Number of clients
N_d	Amount of data uploaded by each client
B	Total bandwidth allocated to a specific HD map layer
T_t	Trigger time of a specific HD map layer
T_w	Waiting time calculated for a specific HD map layer
T_p^j	Preparation time for vehicle j
T_r^j	Response time for vehicle j
μ	Service rate for the queue representing a specific HD map layer pipeline
I	Information value of an HD map tile
W	Lambert W function

customer impatience. However, the approximations of the paper are not suitable for calculating the T_w of each customer after arrival which is essential for our model. A further study [58] explores D/M/1 systems in great detail analyzing every aspect of the network, from costs through probabilities to idle and waiting times. As per this latter independent analysis of D/M/1 queues and the assumptions of our model, T_w can be obtained as $\delta/((1-\delta)\mu)$ with μ being the service rate and δ the smallest absolute root of (2):

$$\Delta = e^{-\mu T_t(1-\Delta)}, \tag{2}$$

where Δ is an internal parameter used for calculating waiting time, variance, and other attributes of D/M/1 queues [58]. This root can be calculated using the Lambert W [59] function:

$$\delta = -\frac{W(-e^{-\mu T_t}\mu T_t)}{\mu T_t} \triangleq -\frac{W_\delta}{\mu T_t}. \tag{3}$$

Consequently, we can get the T_w waiting time of each layer by applying the following formula:

$$T_w = \frac{\delta}{(1-\delta)\mu} = \frac{W_\delta}{\mu^2 T_t(1+\frac{W_\delta}{T_t})}, \tag{4}$$

where W is the Lambert W function.

To limit the possible space of parameters, we introduce the constraints defined in (5). Refer to Table 1 for our notations, where each symbol represents a parameter of a single queue. Constraint (C_1) ensures that there is an upper limit on the number of clients dictated by the size of transmitted data, trigger time and uplink capacity. Ideally, after sending in an HD map data update, every client should receive the merged data (i.e., the up-to-date local HD map) from the FA aggregation node in time, before the next update round. Constraint (C_2) thus makes sure that the waiting time is less than the trigger time. Additionally, constraint (C_3) is required to ensure that the waiting time function returns with a value in the domain of real numbers and it obeys the constraints of D/M/1 queues mentioned in [58]. Otherwise, this value can be in the complex domain because of the utilized Lambert W function. The constraints (C_4), (C_5), and (C_6) keep the μ service rate, the T_t trigger time, and the N_c number of clients within realistic bounds, respectively, as it is unrealistic to have a service rate

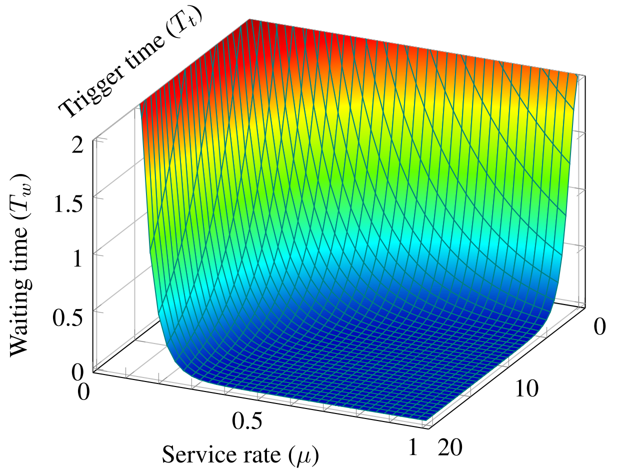

FIGURE 3. Evolution of the T_w waiting time as a function of the μ service rate and the T_t trigger time.

and trigger time under 0 and the number of clients under 1 for an HD map tile that is being updated.

$$\begin{aligned}
&(C_1)\colon N_c N_d \frac{1}{T_t} \leq B;\\
&(C_2)\colon T_w \leq T_t;\\
&(C_3)\colon T_t \mu > 1;\\
&(C_4)\colon \mu > 0;\\
&(C_5)\colon T_t > 0;\\
&(C_6)\colon N_c > 1.
\end{aligned} \tag{5}$$

Figure 3 depicts the change of the T_w waiting time based on the μ service rate and the T_t trigger time within the bounds set by (C_1)–(C_6). As shown by (4), the waiting time is inversely proportional to the trigger time, therefore the logical optimization for the trigger time is to minimize them while obeying the constraints in (5). Later, in Section V, our experiments apply trigger times no smaller than 1 s. This is because our use-case does not require sub-second trigger times to achieve a reasonable information density.

While the T_w waiting time can be determined for the vehicular clients using our queuing model, we also need to take into consideration how much a single client's insights contribute to the aggregated knowledge of the whole community. In our view, the merged insights of every vehicular client providing data for a layer of an HD map tile should convey the following aspects. When few clients contribute data to one of the map tile's layers, these uploads should be considered highly important, as they provide the latest observations, however, only when they can be considered timely and not stale. As the coverage area of an HD map tile gets saturated by clients (consider the case of an intersection that handles high traffic during the morning and afternoon rushes), many of these clients will identify the same objects or environmental conditions, making their updates redundant. Data staleness in this case can originate from saturating even the wireless data communication channels meaning that not every vehicular client can contribute to the map in each iteration, as they might not be able to upload data in time. In this case, insights can be considered stale when trying to be uploaded in subsequent iterations. In our model, we capture these effects using the simplification provided by the I information value transmitted within the FA network. To calculate it, we intuitively define a function that provides values within the easily interpretable range of 0 to 1 in order to aid analysis and general readability. We propose the formulae of (6) that combine two constituent functions to represent the two main moving parts of the system:

$$\begin{aligned}
I &= \frac{7}{3\pi}\arctan(I_1 I_2), \quad \text{where:}\\
I_1 &= 4N_c \frac{1}{\sqrt{600 + {N_c}^2}};\\
I_2 &= \begin{cases} \dfrac{1}{T_w - 1 + T_t}, & \text{if } T_w \leq 1,\\ \dfrac{1}{\ln(T_w + T_t)}, & \text{if } T_w > 1. \end{cases}
\end{aligned} \tag{6}$$

We use the first function, I_1, for taking into consideration the N_c number of communicating clients in the geographic area in focus. The chosen sigmoid-like function can adequately capture the aforementioned phenomenon of diminishing returns. When only a handful of vehicular clients are present, the aggregated data contains less information about the environment, however, this rises quickly as new clients appear. After reaching a certain number of clients, however, the transmitted information becomes highly redundant and thus, the increase of I has to slow down. We chose the constants present in the equation carefully based on several iterations spent on designing the I information value function. The final values adequately flatten the sigmoid-like sub-value of I and avoid its rapid increase when getting close to its limits.

We then utilize the second function, I_2, to determine the information recency. It takes into account the T_w waiting time needed to get the most recent information from the edge nodes. This is also the component where we account for the T_t trigger time.

To make sure that the complete I information value function falls within the desired 0 to 1 range, we use arctan and carefully chosen constants for a scaling transformation. In our opinion, this flattened function, as illustrated in Figure 4 for a fixed 1 s trigger time, can give a good approximation of real-life scenarios.

C. POSSIBLE OPTIMIZATION AVENUES

In order for our system to operate at its peak, we need to find its lowest reaction time (I_2) while amassing the most information that is specific for a given HD map layer (I_1). Tangibly, the first implicates the minimization of the waiting time of each vehicular client, while the second can be represented as the maximization of the information value which we can formalize using the following objective functions:

$$\text{For a given HD map layer } l \text{ of tile } t\text{: } \max\ {}_t^l I \tag{7}$$

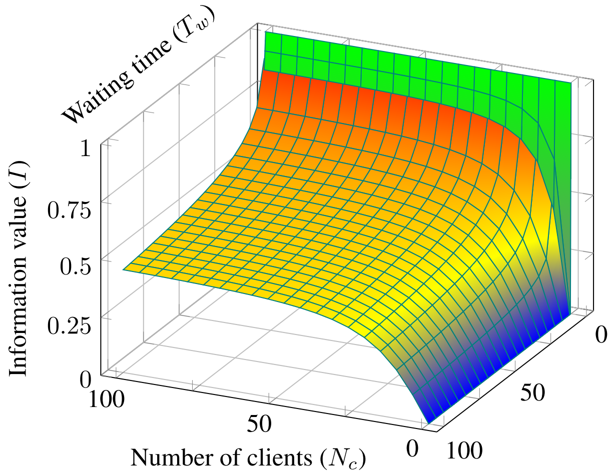

FIGURE 4. Illustration of our chosen I information value function under fixed $T_t = 1$ s trigger time and varying N_c client number and T_w waiting time ($\langle N_c, T_w \rangle \in [0, 100]$) scaled to the [0, 1] region. The two domains of the I_2 component are shaded with different colors: $T_w \leq 1$ is green, $T_w > 1$ is colored blue to orange.

where ${}^l_t I$ is the information value calculated for the specific HD map's layer l and tile t, and each constituent respects the (C_1)–(C_6) constraints of (5).

Due to the proposed I and T_w functions being too complex for analytical optimization techniques, we selected three different numerical optimization methods to find the optimal values. We leveraged the constrained optimization by linear approximation (COBYLA) [60], the Trust Region Constrained Algorithm (TRCA) [61], and the Sequential Least SQuares Programming (SLSQP) [62] optimizer algorithms.

Figure 5 illustrates the paths these algorithms took while reaching the optimal I information value in a configuration with a $\mu = 0.5$ service rate, $B = 1$ Gb/s bandwidth, and $N_d = 1$ MB transferred data size while considering a transient dynamic HD map layer implying a $T_t < 5$ s layer trigger time. As shown in Figure 5 (left), the three different algorithms traversed extremely different paths before settling on similar optimal values (marked in the figure) when considering the T_t trigger time as the input value. Initially, the COBYLA and SLSQP algorithms explore the space via seemingly unpredictable paths. Before finding the $I \approx 0.907$ optimal value, COBYLA already starts to oscillate around $T_t \approx 3.36$ s in the $I > 0.745$ domain. TRCA and SLSQP do not display this oscillation before finding the optimal values of $I \approx 0.907$ also at $T_t \approx 3.36$ s. Contrarily, when the optimization is dependent on the N_c number of clients as the input value, as depicted in Figure 5 (right), the optimization paths are nearly overlapping while the marked optimal values found by the three algorithms retain their slight difference being identical to those found in the previous case. The number of serviceable clients, however, is vastly different, as $N_c^{\text{COBYLA}} \approx 52.42$, $N_c^{\text{TRCA}} \approx 1293.28$ and $N_c^{\text{SLSQP}} \approx 2705.86$ which can be attributed to the different inner mechanisms of the explored algorithms.

We draw two main conclusions from the results of multiple different optimization runs with different configurations (not only the illustrative example depicted in Figure 5). First, the changes in the T_w waiting times are almost negligible in the range needed for optimal HD map transfer. Assuming instantaneous uplink and downlink transfers, this can be attributed to the properties of D/M/1 queues. Second, we can draw insights into how the relation of the B bandwidth and the I information value can lead to an optimal FA network. When the bandwidth drops, the N_c number of serviceable clients drastically decreases too, however, the I information value can still be retained by dynamically decreasing the T_t trigger time of the clients contributing to an HD map layer. This is especially important for highly dynamic layers, where even a small number of clients can provide a reasonable amount of information if they are triggered frequently using a low T_t trigger time.

Using a combination of COBYLA, TRCA, and SLSQP optimizes our approach by ensuring we are equipped for a variety of problem conditions. COBYLA is excellent for dealing with non-differentiable regions, TRCA adeptly handles complex constraints with its trust-region methodology, and SLSQP exploits smooth regions efficiently, offering a comprehensive exploration of the optimization landscape, maximizing the potential to avoid local minima and gain valuable insights about the problem.

V. NUMERICAL EVALUATION

In order to reinforce our theoretical model of the proposed federated HD map aggregating network, in this section, we present empirical results gained by running simulation scenarios over a vast range of model parameters. For these to carry indications of the prospective behavior of the proposed solution in a production environment, we employ a complex testbed built on top of an industry-recognized simulation framework. Thus, we first thoroughly discuss the high-level implementation of our testbed, then we proceed by defining the context and the metainformation for our simulation scenarios. Finally, we discuss our results and draw conclusions about the practical applications of our system.

A. TESTBED DESCRIPTION

Our testbed consists of a custom-written application simulator that relies on the Simulation of Urban MObility (SUMO) framework [63]. In recent years, SUMO has become an industry-leading solution for virtual vehicle simulation due to its ability to easily handle large traffic networks, while holistically managing every prospective actor, ranging from automobiles and pedestrians to traffic lights and roadside equipment. Connecting this framework to our defined, application-level testbed was made possible by exploiting the TraCI interface [64]. This provides an interface through which our simulator can gain access to every volatile data regarding the test environment, most importantly the relative position of the vehicles compared to the defined edge node.

B. PROOF-OF-CONCEPT APPLICATION SIMULATOR

At the implementation level, as the simulation in SUMO gradually unfolds, we perform the operations required for

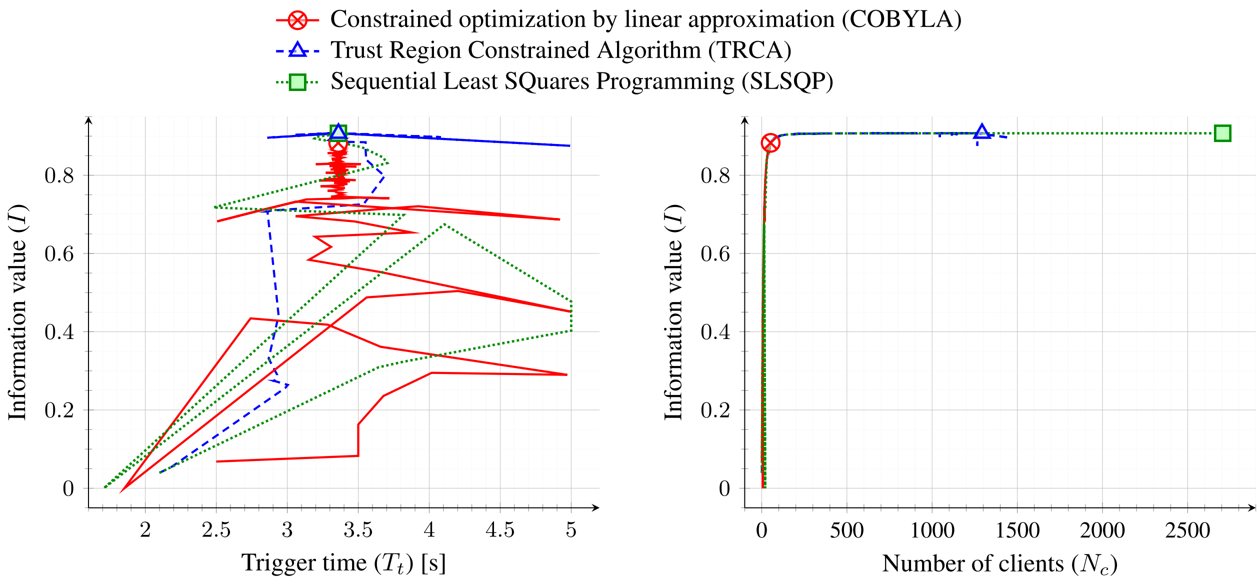

FIGURE 5. Paths taken by various algorithms for determining the optimal I information value as a function of the T_t trigger time (left) and N_c number of clients (right). End results are marked.

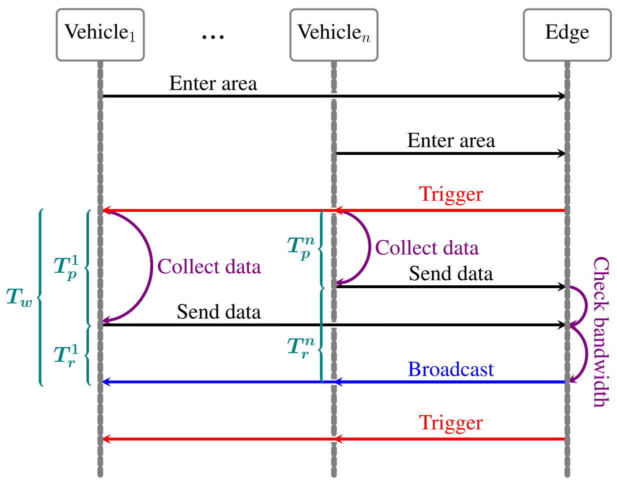

FIGURE 6. System flow of federated analysis of HD map data.

the federated HD-map analytics in the same fashion. Every member of each actor group has a step listener which executes data-driven communications between the actors when prompted by the TraCI interface. In our proof-of-concept application, we only consider vehicles and edge nodes, but these can be optionally expanded.

During the initialization of the SUMO simulation, we create a Python object for a single edge node and get its step listener registered into the event cycle through the TraCI interface. This process is also performed for each vehicle that appears in the simulation throughout the entire course of the simulation period. After their registration, the step listeners get invoked at every discrete timeslot when a predefined step-length duration of time has passed, performing their actions periodically, based on the current state of the Python objects representing them.

As depicted in Figure 6, the edge node regularly triggers vehicles that are residing in its transmission area. Following our terminology, this happens every time after a T_t trigger time has passed. Upon this trigger stimulus, the vehicles collect, package, and then send data regarding their immediate surroundings to update a specific HD map layer. In accordance with (1), this period is labeled as the T_p preparation time. The vehicles then wait for each other to finish their data preparation phase, enabling the edge node to perform the data aggregation via FA. When this is complete, all vehicles present in the transmission area by this time receive the aggregated information as the latest version of the respective HD map layer. For those vehicles that have participated in the data collection, this period is referred to as the T_r response time, as defined by (1). We note that $T_p + T_r = T_w$ is the same for all vehicles participating in the federated analytics process, both in our model and during the simulation. Finally, following the aggregation, until the current trigger period lasts, all participants are idle, as they wait for the subsequent trigger stimulus.

We consider the capabilities of 5G networks indirectly, in terms of current and prospective bandwidth. Being an integral part of our theoretical model, we aim at emulating the network bandwidth precisely. Every time a new vehicle starts its data upload, the simulator calculates whether there is enough bandwidth remaining in the current trigger period for the process to complete. More precisely, the simulator checks whether the $B_r\tau$ product of the remaining bandwidth and upload time during the trigger period is still able to accommodate the data with the given size to finish uploading. This constraint is the most restrictive in cases when having short trigger periods of < 1 s. However, considering our use case, such low trigger times are not sensible, and we simulate cases only with $T_t > 1$ s where this restriction does not cause issues.

Note that for the sake of simplicity, here we described the system with a single edge node, although, in a practical

TABLE 2. Attributes and their investigated values in the simulated scenarios.

Attribute name	Set of values
d: Vehicle density	200/km^2
$\mathbb{E}[N_d]$: Average data size	300 kB
$\mathbb{E}[T_p]$: Average preparation time	$T_t \cdot 0.1$ s
T_t: Trigger time	1, 10, 100 s
D: Edge diameter	0.5, 1, 2, 4 km
B: 5G uplink bandwidth	10, 100, 1000 Mb/s

example, multiple edge nodes may be placed to cover a larger area, e.g., cities or highways. Moreover, each edge node should be able to maintain the aggregation of every HD map layer separately. Our proposed system can incorporate these refined requirements simply by transmitting newly aggregated map layers between neighboring edge nodes and by defining distinct trigger periods for every HD map layer. We also note that our implementation can be extended with a more precise 5G channel model incorporating signal reflection, interference, and other low-level network properties, however, we leave this aspect for future work.

C. RESULTS OF THE NUMERICAL PERFORMANCE EVALUATION

Building on the parameters discussed in our mathematical model description of Section IV, we created multiple simulation scenarios to determine their effects on our system. We reuse the length of the T_t trigger time and the exponential distribution for the T_p preparation time discussed there. We also leverage the actual density of vehicular clients in the transmission area which is influenced by an amalgamation of multiple factors: the overall vehicle density in the input map, the diameter of the edge node's coverage area, and the 5G uplink bandwidth available to the clients. We give a summary of all the factors used in our scenarios in Table 2. Value selection for the used parameters was motivated to make a comprehensive study by taking essential factors into account with a broad-ranged value set. Within the scope of our proposed use-case, values were selected to remain meaningful and down-to-earth, but also to fulfill our final target of widespread dimensioning of our proposed system. Thus some values follow a logarithmic scale, to grasp a wide enough scale.

We base our scenarios on a custom SUMO map simulating a dense metropolitan area with a constant vehicle load. To provide an adequate stress test for cases where traffic jams might frequently occur, we set this load to $d = 200$ vehicles/km^2. This is a reasonable but slightly elevated density compared to related studies [65], [66] that suggest an order of magnitude of 100 vehicles/km^2 for this parameter. Keeping the execution time of each simulated scenario under reins further limited our selection of this value. Within this setting, our simulator randomly generates vehicle route patterns to produce the targeted average vehicle density. We set an $N_d = 300$ kB average transmitted data size by assuming the exchange of privacy-preserving FA insights between clients and the edge node performing the FA aggregation.

We took the Cartesian product of the value sets listed in Table 2 so that one particular combination, e.g., a tuple of $\langle D = 0.5\,\text{km}, T_t = 1\,\text{s}, B = 10\,\text{Mb/s}\rangle$, defined the input values for a given simulation scenario. We repeated every scenario 10 times and averaged their individual results. In each case, we collected the observed changes caused in the target variables of the T_w average waiting time of every vehicle within the trigger period and the I information value which is the numerical output of the information function as per defined in Section IV. In the following, we analyze the observed effects.

1) DIAMETER

In Figure 7, we depict how different edge node coverage area (map tile) diameters affect the waiting time and the information value. First, let us observe Figure 7a: in every test scenario, the waiting time converges to the trigger time as the diameter grows. Given that more vehicles become reachable, the uplink load becomes greater up to reaching the set bandwidth limit. This is more conspicuous if we check the lines corresponding to lower trigger times and high, Gb/s 5G uplink bandwidth. These approach the trigger time at a much slower rate as they can accommodate more traffic. On the other hand, the impact of the trigger time is almost equally large. As the trigger time becomes greater, the effect of the uplink bandwidth gradually becomes less apparent on the waiting time. Figure 7b depicts the relation between the diameter and the information value: we can discover a similar phenomenon here as well. The reason is again the growing vehicle density in the transmission area.

The optimal diameter can be found at approximately $D = 1$ km in most cases if such dense areas are considered. The information value reaches its peak at this point while the waiting time does not deteriorate much compared to the scenario with a $D = 0.5$ km diameter. If the observed area has lower vehicle density then a bigger diameter can provide a better quality of service: low waiting time and high information value. This is due to the fact that if the vehicle density is not too high in a large cell then the clients are able to finish their upload before the trigger period ends (i.e., the uplink bandwidth capacity is able to accommodate all clients). This causes both the waiting time and the information value to change for the better. Conversely, when a high vehicle density is considered then the same phenomenon can be observed for large diameters as for low cell diameters depicted in Figures 7a–b.

2) TRIGGER TIME

Figure 8 shows the effects of changes in the trigger time induce in both the waiting time and the information value. In the case of the former (Figure 8a), every test scenario is strikingly linear on the log-log scale. This is mainly due to the fixed connection between the expected value of

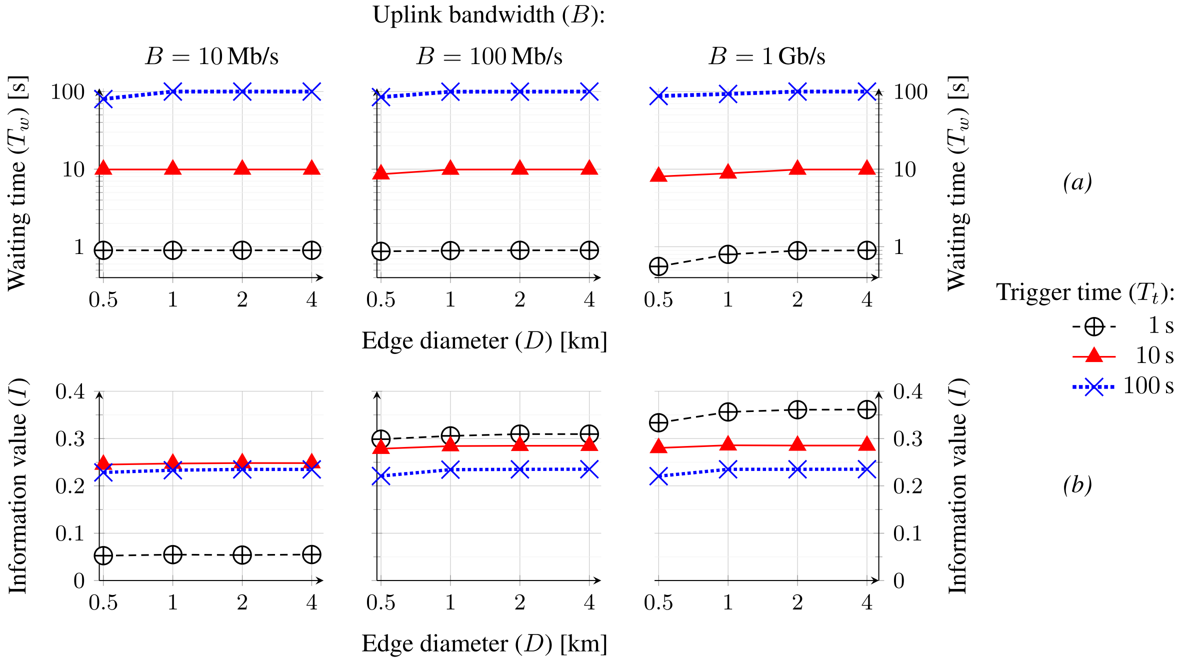

FIGURE 7. Effect of the D edge diameter on the T_w waiting time (a row, top) and I information value (b row, bottom). Horizontal axes have a base-2 logarithmic scale, subfigures of the top row have vertical axes with a base-10 logarithmic scale.

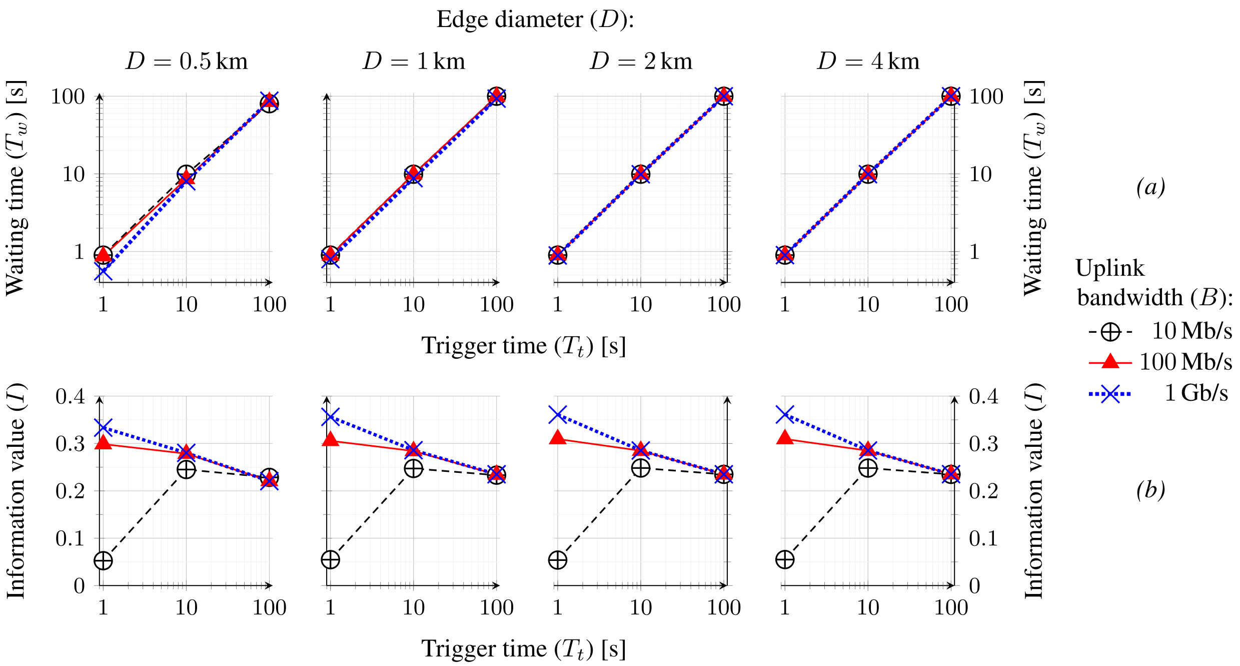

FIGURE 8. Effect of the T_t trigger time on the T_w waiting time (a row, top) and I information value (b row, bottom). Horizontal axes have a base-10 logarithmic scale, subfigures of the top row have vertical axes with a base-10 logarithmic scale.

the preparation time and the trigger time in our simulation input (see Table 2). In terms of steepness, there are slight differences among the lines: when expanding the available uplink bandwidth, the clients get a higher chance in each trigger period to finish data collection before the new trigger period starts. Thus, they are more likely to attain a lower response time (latency), consequently lowering the waiting time. This phenomenon is true until the vehicle density allows it. As soon as the area becomes overpopulated compared to what the uplink capacity can carry, the response time also starts to grow, since the vehicles are not able to finish their uploads soon enough anymore. Hence, the respective lines become steeper in the figure.

When considering the information value in Figure 8b, let's separate the plots with medium or high bandwidth from the ones with low bandwidth and examine them independently.

Upon observing the first group, it is important to notice that they decline as the trigger period widens. However, the

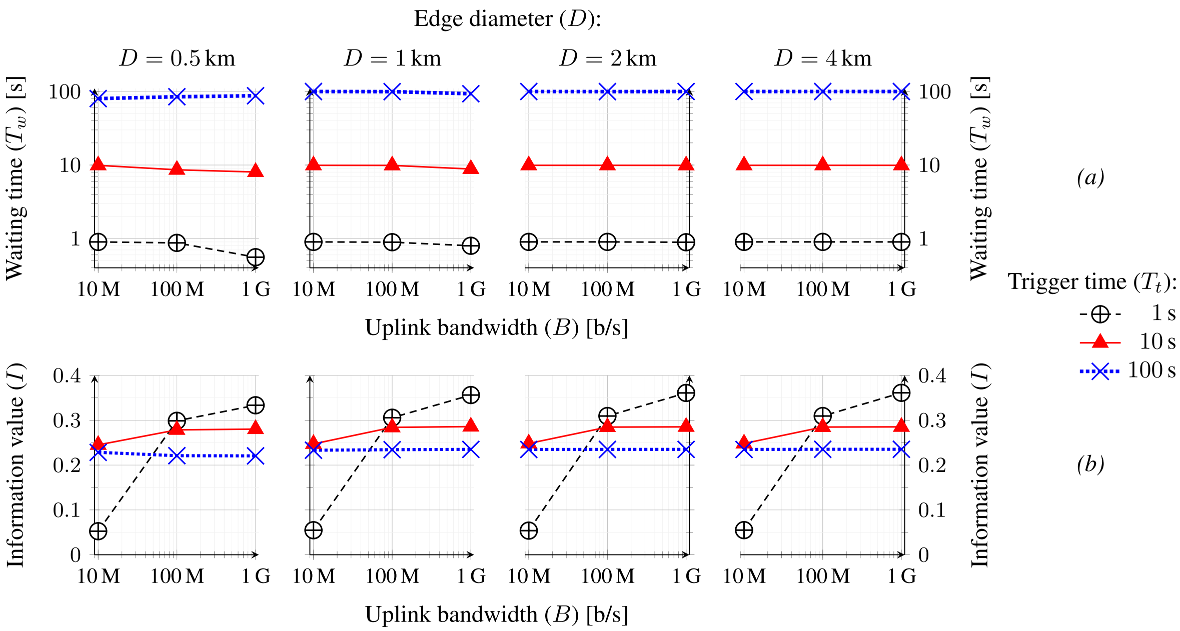

FIGURE 9. Effect of the B uplink bandwidth on the T_w waiting time (a row, top) and I information value (b row, bottom). Horizontal axes have a base-10 logarithmic scale, subfigures of the top row have vertical axes with a base-10 logarithmic scale.

convexity/concavity of the function varies across the test scenarios. This feature is mainly dependent on the uplink bandwidth. If it is in the 100 Mb/s region, the function is concave, conversely, if it reaches the volume of Gb/s, the function becomes convex. Regardless of the shape, each function has its maximum at the beginning of the scale, at around 1 s. This is a direct cause of the fact that we considered only urban areas with a high vehicle density. When these circumstances are given, it is advisable to maintain a low trigger time because then the information value is at its highest. Other factors, on the other hand, such as the characteristics of a specific HD map layer, the incurred transmission pollution, or other cost-related insights should also be taken into account when setting the trigger time.

Moving on to the second observation group, the features are highly different. Every function stays concave, reaching their maximum values only at approximately $T_t = 10$ s. As a result, the optimal setting differs from the previously stated, i.e., when the available bandwidth is low, it is advisable to make use of larger trigger times.

3) BANDWIDTH

Similarly to prior plots, we depict the effect of the available uplink bandwidth on the waiting time and information value. Figure 9 displays some similarities to the diameter-related trends shown in Figure 7, especially considering the waiting time metric. T_w shows a great dependency on the uplink bandwidth, stemming from the fact that it is one of the two main contributors that determine whether vehicles can achieve a low T_p, or even finish, in a specific trigger period. As the available uplink bandwidth spreads out, so does the waiting time become increasingly lower across all scenarios, however, more so in cases with a smaller diameter.

The information value plot in Figure 9b has a similar duality as shown in Figure 8b. In this case, however, the length of the trigger period proves to be the distinctive factor. The higher the T_t trigger time is, the flatter each function gets, culminating in nearly stagnating lines when T_t reaches its maximum. In terms of diameter, an opposite relationship can be observed: as the diameter grows for scenarios corresponding to each distinct trigger time value, the lines are becoming increasingly steeper.

4) PREPARATION TIME

As with the rest of the plots, we depict the effect that the change in the expected value of the preparation time causes in the waiting time and information value. Figure 10 displays some similarities to the diameter-related trends, especially in waiting time metrics. The $\mathbb{E}[T_p]$ is an immense factor in T_w, stemming from the definition of the latter. Hence, if T_p grows then the probability of getting a value much closer to T_t also increases because the rate of the exponential distribution gets lower. Conclusively, as Figure 10a unmistakably illustrates, this results in a rising T_w.

The information value plot has an interesting duality created by the available uplink bandwidth in each scenario. Figure 10b depicts that a low bandwidth (100 Mb/s) induces a deteriorating trend in the information value, while the gigabit bandwidth plots show increasing or stagnating trends, as $\mathbb{E}[T_p]$ enlarges. The explanation revolves around the ability of each vehicle to finish its data upload as early as possible, given a specific T_p. If the remaining bandwidth does not further constrain the upload, then the information value is indifferent toward this. On the other hand, if the remaining bandwidth has become critically low for slow clients with a large T_p, it severely constrains most vehicles. In this case,

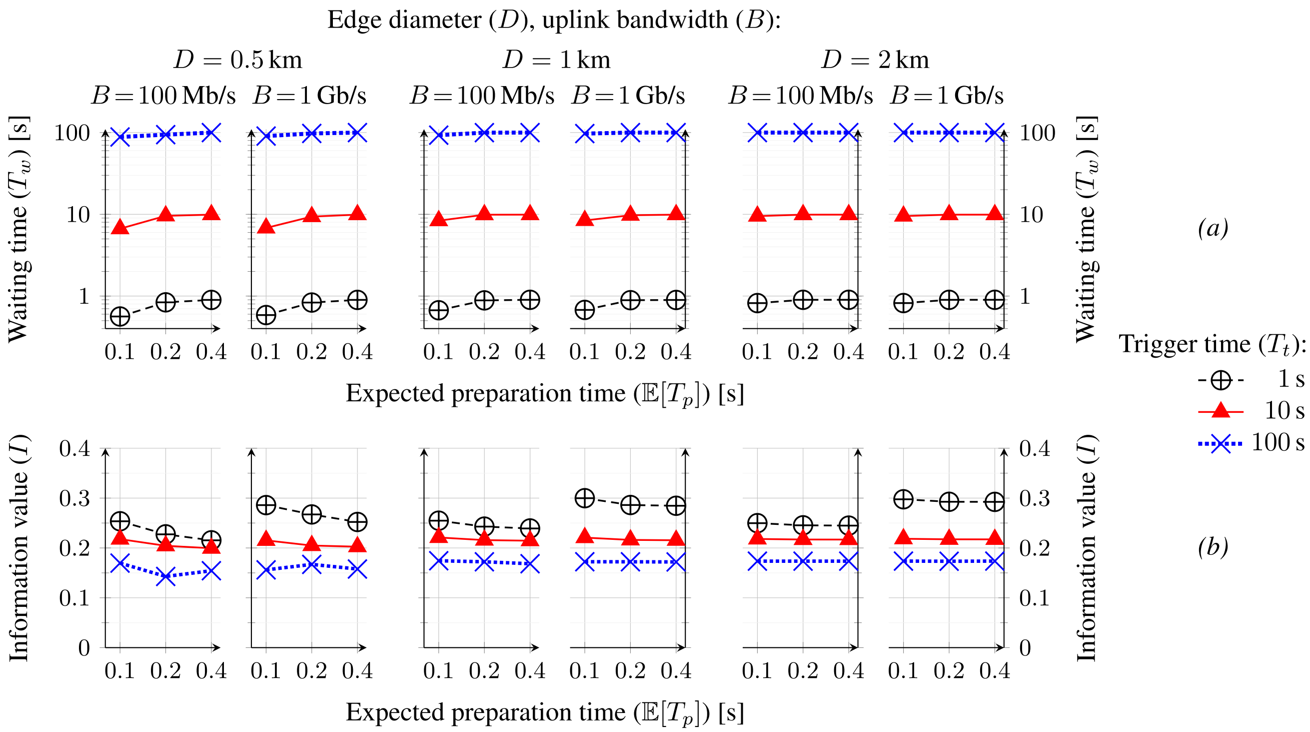

FIGURE 10. Effect of the T_p preparation time on the T_w waiting time (*a* row, top) and I information value (*b* row, bottom). Horizontal axes have a base-2 logarithmic scale, subfigures of the top row have vertical axes with a base-10 logarithmic scale.

only a few can upload within their T_t, thus causing the information value to drop.

In summary, we can say that by employing numerical simulations, we gathered a vast set of evidence for determining the properties of our theoretical model. We tightened the input and constraint variables to a defined set of values and calculated the results based on these. Our findings show that in areas with a high vehicle density, such as metropolitan centers, the best setting for both end-to-end latency and information value would be to have HD maps tiles and corresponding 5G cells with a relatively small transmission area (approximately 1 km in diameter) and as low trigger time as it is sensible for the specific use-case. Additionally, in order to reach as low a preparation time as possible, having high bandwidth capabilities in such areas is crucial. These requirements become more tangible than ever. The growing 5G coverage around the globe and increasing upload speeds bring us closer to such scenarios. Although the required Gb/s uplink speeds are not yet available publicly in dense areas and high mobility scenarios, laboratory experiments have demonstrated such speeds with 5G test hardware. As these transition to publicly available environments, along with the significant developments that are being carried out in the field of intelligent transportation, we expect that similar applications will emerge in the coming years.

VI. CONCLUSION

In this work, we have introduced our proposal for creating an HD maps information aggregation and distribution system that leverages the concept of federated analytics and 5G edge infrastructure under the hood. Our system is designed to be fast and economical when dealing with tasks related to the construction of multi-layered maps—containing static and dynamic traffic and contextual data—in a crowd-sourced fashion. Both FA and 5G proved to be crucial in achieving our goals. FA provides privacy-preserving features: it enables local data preprocessing and a shared model can be built by distributing only derived insights. This is beneficial for sparing bandwidth and reducing latency as well. 5G also reduces latency via its efficient wireless links and edge computing capabilities. Our paper includes both theoretical and experimental analyses aimed at optimizing HD map update scheduling and the grouping of clients for creating the HD map of a specific area.

Our model is built on D/M/1 queues which brings forth deterministic HD map update periods. While this can be considered a limitation, we believe it to be an advantage in two ways. First, it makes possible to express the end-to-end latency of the FA process in closed formulae. Second, it can serve as an enabler for the easier adaptation of different clients to the FA merge process. A further limitation is introduced by our definition of the I information value which conveys how valuable the contribution of a vehicular client is on the level of our model. While I displays diminishing returns at high map tile saturation, it lacks configurable parameters with which the fine-grained quality or fidelity of the observations provided by the modeled clients can be modeled. We also limit our investigation to high-density urban scenarios (traffic density, HD map tile size) in order to observe its performance under stressing situations.

We identify these as future improvement possibilities where the effects of a more fine-tuned information value function could be investigated under different (low-density urban, rural) scenarios.

Despite these limitations, our findings indicate that our proposed system is viable and can deliver relevant information to end users within an acceptable timeframe. Specifically, our results suggest the use of 5G small cells with a diameter of approximately 1 kilometer, offering high Gb/s uplink speeds and selecting low but sensible trigger times to achieve optimal waiting times and information values.

As our measurements were taken in a simulated environment, in the future we plan to broaden the investigations to offer a more comprehensive view on the effects of various (environmental and infrastructural) factors for paving the way for practical applications. We plan to leverage the SLICES (Scientific Large-Scale Infrastructure for Computing/Communication Experimental Studies) [67] project's infrastructure. It aims to support experimental research on various aspects of artificial intelligence, cloud/edge computing, and next-generation (5G/6G) telecommunications networks that is also perfectly in line with the scope of our work.

REFERENCES

[1] Continental AG. *Continental Continues to Drive Forward the Development of Server-based Vehicle Architectures*. Accessed: Sep. 7, 2023. [Online]. Available: https://www.continental.com/en/press/press-releases/20210728-cross-domain-hpc/

[2] F. Chow. (2022). *The New Standard: Red Hat in-Vehicle Operating System in Modern and Future Vehicles*. Accessed: Sep. 7, 2023. [Online]. Available: https://www.redhat.com/en/blog/new-standard-red-hat-vehicle-operating-system-modern-and-future-vehicles

[3] J. Marquez-Barja, B. Lannoo, D. Naudts, B. Braem, B. C. Donato, V. Maglogiannis, S. Mercelis, R. Berkvens, P. Hellinckx, M. Weyn, I. Moerman, and L. Steven, "Smart highway: ITS-G5 and C2VX based testbed for vehicular communications in real environments enhanced by edge/cloud technologies," in *Proc. Eur. Conf. Netw. Commun. (EuCNC), Abstr.*, 2019, p. 2.

[4] Amazon Web Services. (2022). *5G Edge Computing Infrastructure—AWS Wavelength*. Accessed: Sep. 7, 2023. [Online]. Available: https://aws.amazon.com/wavelength/

[5] Y. Khalidi. (Mar. 2020). *Microsoft Partners With the Industry to Unlock New 5G Scenarios With Azure Edge Zones*. Accessed: Sep. 7, 2023. [Online]. Available: https://azure.microsoft.com/hu-hu/blog/microsoft-partners-with-the-industry-to-unlock-new-5g-scenarios-with-azure-edge-zones/

[6] A. Phadke. (Dec. 2020). *Bringing Partner Applications to the Edge With Google Cloud*. Accessed: Sep. 7, 2023. [Online]. Available: https://cloud.google.com/blog/topics/anthos/anthos-for-telecom-puts-google-cloud-partners-apps-at-the-edge

[7] GA Association, "White paper: C-V2X use cases and service level requirements volume I," 5GAA Automot. Assoc., München, Germany, Tech. Rep. T-200111, Dec. 2020. Accessed: Jul. 9, 2023.

[8] D. Ramage and S. Mazzocchi. (2020). *Federated Analytics: Collaborative Data Science Without Data Collection*. Accessed: Jul. 9, 2023. [Online]. Available: https://ai.googleblog.com/2020/05/federated-analytics-collaborative-data.html

[9] L. Toka, M. Konrád, I. Pelle, B. Sonkoly, M. Szabó, B. Sharma, S. Kumar, M. Annavazzala, S. T. Deekshitula, and A. Antony Franklin, "5G on the roads: Optimizing the latency of federated analysis in vehicular edge networks," in *Proc. IEEE/IFIP Netw. Operations Manage. Symp.*, May 2023, pp. 1–5.

[10] J. Xie, J. Tang, and S. Liu, "An energy-efficient high definition map data distribution mechanism for autonomous driving," 2020, *arXiv:2010.05233*.

[11] X. Wu, X. Wen, L. Wang, W. Zheng, Z. Lu, and L. Liu, "A cooperated approach between V2I and V2V for high definition map dissemination in automated driving," in *Proc. 33rd Gen. Assem. Scientific Symp. Int. Union Radio Sci.*, Aug. 2020, pp. 1–4.

[12] X. Xu, S. Gao, and M. Tao, "Distributed online caching for high-definition maps in autonomous driving systems," *IEEE Wireless Commun. Lett.*, vol. 10, no. 7, pp. 1390–1394, Jul. 2021.

[13] H. Peng, Q. Ye, and X. S. Shen, "SDN-based resource management for autonomous vehicular networks: A multi-access edge computing approach," *IEEE Wireless Commun.*, vol. 26, no. 4, pp. 156–162, Aug. 2019.

[14] R. Zhang and K. Cai, "The application of edge computing in high-definition maps distribution," in *Proc. The 2nd World Symp. Softw. Eng.*, Sep. 2020, pp. 116–121.

[15] V. P. Chellapandi, L. Yuan, S. H. Zak, and Z. Wang, "A survey of federated learning for connected and automated vehicles," 2023.

[16] K. Doshi and Y. Yilmaz, "Federated learning-based driver activity recognition for edge devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 3337–3345.

[17] C. Zhao, Z. Gao, Q. Wang, K. Xiao, Z. Mo, and M. J. Deen, "FedSup: A communication-efficient federated learning fatigue driving behaviors supervision approach," *Future Gener. Comput. Syst.*, vol. 138, pp. 52–60, Jan. 2023.

[18] H. Zhang, J. Bosch, and H. H. Olsson, "End-to-end federated learning for autonomous driving vehicles," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2021, pp. 1–8.

[19] M. P. Aparna, R. Gandhiraj, and M. Panda, "Steering angle prediction for autonomous driving using federated learning: The impact of vehicle-to-everything communication," in *Proc. 12th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Jul. 2021, pp. 1–7.

[20] C. Wang, X. Chen, J. Wang, and H. Wang, "ATPFL: Automatic trajectory prediction model design under federated learning framework," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 6553–6562.

[21] C. Koetsier, J. Fiosina, J. N. Gremmel, J. P. Müller, D. M. Woisetschläger, and M. Sester, "Detection of anomalous vehicle trajectories using federated learning," *ISPRS Open J. Photogramm. Remote Sens.*, vol. 4, Apr. 2022, Art. no. 100013.

[22] M. Han, K. Xu, S. Ma, A. Li, and H. Jiang, "Federated learning-based trajectory prediction model with privacy preserving for intelligent vehicle," *Int. J. Intell. Syst.*, vol. 37, no. 12, pp. 10861–10879, Dec. 2022.

[23] Y. Fu, C. Li, F. R. Yu, T. H. Luan, and Y. Zhang, "A selective federated reinforcement learning strategy for autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 2, pp. 1655–1668, Feb. 2023.

[24] A. M. Elbir, B. Soner, S. Çöleri, D. Gündüz, and M. Bennis, "Federated learning in vehicular networks," in *Proc. IEEE Int. Medit. Conf. Commun. Netw. (MeditCom)*, Sep. 2022, pp. 72–77.

[25] Y. Liu, A. Huang, Y. Luo, H. Huang, Y. Liu, Y. Chen, L. Feng, T. Chen, H. Yu, and Q. Yang, "Fedvision: An online visual object detection platform powered by federated learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, pp. 13172–13179, 2020.

[26] P. Yu and Y. Liu, "Federated object detection: Optimizing object detection model with federated learning," in *Proc. 3rd Int. Conf. Vis., Image Signal Process.* New York, NY, USA: Association for Computing Machinery, 2020, pp. 1–6.

[27] L. Li, Y. Fan, M. Tse, and K.-Y. Lin, "A review of applications in federated learning," *Comput. Ind. Eng.*, vol. 149, Nov. 2020, Art. no. 106854.

[28] D. Jallepalli, N. C. Ravikumar, P. V. Badarinath, S. Uchil, and M. A. Suresh, "Federated learning for object detection in autonomous vehicles," in *Proc. IEEE 7th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Aug. 2021, pp. 107–114.

[29] G. Rjoub, O. A. Wahab, J. Bentahar, and A. S. Bataineh, "Improving autonomous vehicles safety in snow weather using federated YOLO CNN learning," in *Mobile Web and Intelligent Information Systems*. Cham, Switzerland: Springer, 2021, pp. 121–134.

[30] K. D. Stergiou, K. E. Psannis, V. Vitsas, and Y. Ishibashi, "A federated learning approach for enhancing autonomous vehicles image recognition," in *Proc. 4th Int. Conf. Comput. Commun. Internet (ICCCI)*, Jul. 2022, pp. 87–90.

[31] S. B. Prathiba, G. Raja, S. Anbalagan, K. Dev, S. Gurumoorthy, and A. P. Sankaran, "Federated learning empowered computation offloading and resource management in 6G-V2X," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3234–3243, Sep. 2022.

[32] F. Eckermann, M. Kahlert, and C. Wietfeld, "Performance analysis of C-V2X mode 4 communication introducing an open-source C-V2X simulator," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–5.
[33] Z. Du, C. Wu, T. Yoshinaga, K. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular Internet of Things: Recent advances and open issues," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 45–61, 2020.
[34] L. Liu, J. Zhang, S. H. Song, and K. B. Letaief, "Client-edge-cloud hierarchical federated learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2020, pp. 1–6.
[35] L. Toka, M. Konrád, I. Pelle, B. Sonkoly, M. Szabó, B. Sharma, S. Kumar, M. Annavazzala, S. T. Deekshitula, and A. Antony Franklin, "Federated learning for vehicular coordination use cases," in *Proc. IEEE/IFIP Netw. Operations Manage. Symp.*, May 2023, pp. 1–6.
[36] B. Sonkoly, J. Czentye, M. Szalay, B. Németh, and L. Toka, "Survey on placement methods in the edge and beyond," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2590–2629, 4th Quart., 2021.
[37] A. Moubayed, A. Shami, P. Heidari, A. Larabi, and R. Brunner, "Edge-enabled V2X service placement for intelligent transportation systems," *IEEE Trans. Mobile Comput.*, vol. 20, no. 4, pp. 1380–1392, Apr. 2021.
[38] I. Pelle, F. Paolucci, B. Sonkoly, and F. Cugini, "P4-assisted seamless migration of serverless applications towards the edge continuum," *Future Gener. Comput. Syst.*, vol. 146, pp. 122–138, Sep. 2023.
[39] Z. Wang, D. Zhao, M. Ni, L. Li, and C. Li, "Collaborative mobile computation offloading to vehicle-based cloudlets," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 768–781, Jan. 2021.
[40] X. Dai, Z. Xiao, H. Jiang, H. Chen, G. Min, S. Dustdar, and J. Cao, "A learning-based approach for vehicle-to-vehicle computation offloading," *IEEE Internet Things J.*, vol. 10, no. 8, pp. 7244–7258, Apr. 2023.
[41] Y. Chen, F. Zhao, X. Chen, and Y. Wu, "Efficient multi-vehicle task offloading for mobile edge computing in 6G networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 4584–4595, May 2022.
[42] Y. Li, X. Tao, X. Zhang, J. Liu, and J. Xu, "Privacy-preserved federated learning for autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8423–8434, Jul. 2022.
[43] Y. Wang, L. Xiong, X. Niu, Y. Wang, and D. Liang, "A federated learning based privacy-preserving data sharing scheme for Internet of Vehicles," in *Frontiers in Cyber Security*. Singapore: Springer Nature, 2022, pp. 18–33.
[44] J. Casaletto, A. Bernier, R. McDougall, and M. S. Cline, "Federated analysis for privacy-preserving data sharing: A technical and legal primer," *Annu. Rev. Genomics Human Genet.*, vol. 24, no. 1, pp. 1–5, 2023.
[45] X. Yin, Y. Zhu, and J. Hu, "A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions," *ACM Comput. Surveys*, vol. 54, no. 6, pp. 1–36, Jul. 2021.
[46] R. Liu, J. Wang, and B. Zhang, "High definition map for automated driving: Overview and analysis," *J. Navigat.*, vol. 73, no. 2, pp. 324–341, Mar. 2020.
[47] Ericsson. (2020). *5G Evolution: 3GPP Releases 16 & 17 Overview*. Accessed: Jul. 9, 2023. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/ericsson-technology-review/articles/5g-nr-evolution
[48] T. Theodoropoulos, A. Makris, A. Boudi, T. Taleb, U. Herzog, L. Rosa, L. Cordeiro, K. Tserpes, E. Spatafora, A. Romussi, E. Zschau, M. Kamarianakis, A. Protopsaltis, G. Papagiannakis, and P. Dazzi, "Cloud-based XR services: A survey on relevant challenges and enabling technologies," *J. Netw. Netw. Appl.*, vol. 2, no. 1, pp. 1–22, 2022.
[49] B. Sonkoly, B. G. Nagy, J. Dóka, Z. Kecskés-Solymosi, J. Czentye, B. Formanek, D. Jocha, and B. P. Gero, "Towards an edge cloud based coordination platform for multi-user AR applications built on open-source SLAMs," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops (VRW)*, Mar. 2023, pp. 923–924.
[50] H. Wu, H. Zhang, J. Cheng, J. Guo, and W. Chen, "Perspectives on point cloud-based 3D scene modeling and XR presentation within the cloud-edge-client architecture," *Vis. Informat.*, Jul. 2023.
[51] F. Alriksson, D. H. Kang, C. Phillips, J. L. Pradas, and A. Zaidi, "XR and 5G: Extended reality at scale with time-critical communication," *Ericsson Technol. Rev.*, vol. 2021, no. 8, pp. 2–13, Aug. 2021.
[52] J. Dóka, B. G. Nagy, M. A. U. Rehman, D.-H. Kim, B.-S. Kim, L. Toka, and B. Sonkoly, "Ar over NDN: Augmented reality applications and the rise of information centric networking," in *Proc. SIGCOMM Poster Demo Sessions*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 44–45.
[53] Z. Yang, M. Chen, K.-K. Wong, H. V. Poor, and S. Cui, "Federated learning for 6G: Applications, challenges, and opportunities," *Engineering*, vol. 8, pp. 33–41, Jan. 2022.
[54] M. Sugimoto, *Cloud XR (Extended Reality: Virtual Reality, Augmented Reality, Mixed Reality) and 5G Mobile Communication System for Medical Image-Guided Holographic Surgery and Telemedicine*. Singapore: Springer, 2022, pp. 381–387.
[55] B. G. Nagy, J. Dóka, S. Rácz, G. Szabó, I. Pelle, J. Czentye, L. Toka, and B. Sonkoly, "Towards human–robot collaboration: An industry 4.0 VR platform with clouds under the hood," in *Proc. IEEE 27th Int. Conf. Netw. Protocols (ICNP)*, Oct. 2019, pp. 1–2.
[56] O. Alnajar and A. Barnawi, "Tactile Internet of Federated things: Toward fine-grained design of FL-based architecture to meet TIoT demands," *Comput. Netw.*, vol. 231, Jul. 2023, Art. no. 109712.
[57] G. P. Cosmetatos and G. P. Prastacos, "An approximate analysis of the D/M/1 queue with deterministic customer impatience," *RAIRO Oper. Res.*, vol. 19, no. 2, pp. 133–142, 1985.
[58] B. Jansson, "Choosing a good appointment system—A study of queues of the type (D, M, 1)," *Oper. Res.*, vol. 14, no. 2, pp. 292–312, Apr. 1966.
[59] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, no. 1, pp. 329–359, 1996.
[60] S. Gomez and J. Hennart, *Advances in Optimization and Numerical Analysis*. (Mathematics and Its Applications). Amsterdam, The Netherlands: Springer, 2013.
[61] M. Ulbrich and S. Ulbrich, "Non-monotone trust region methods for nonlinear equality constrained optimization without a penalty function," *Math. Program.*, vol. 95, no. 1, pp. 103–135, Jan. 2003.
[62] D. Kraft, "A software package for sequential quadratic programming," Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt, DLR German Aerosp. Center Inst. for Flight Mech., Cologne, Germany, Tech. Rep. DFVLR-FB 88-28, 1988.
[63] German Aerospace Center. (2022). *SUMO Documentation*. Accessed: Jul. 9, 2023. [Online]. Available: https://sumo.dlr.de/docs/index.html
[64] German Aerospace Center. (2022). *TraCI*. Accessed: Jul. 9, 2023. [Online]. Available: https://sumo.dlr.de/docs/TraCI.html
[65] A. Loder, L. Ambühl, and M. Menendez, "Understanding traffic capacity of urban networks," *Sci. Rep.*, vol. 9, p. 16283, Nov. 2019.
[66] H. B. Banaag, M. S. Litana, and R. V. Ramos, "Vehicle density estimation in quezon city using object-based feature extraction on satellite images," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vols. XLVI-4/W6-2021, pp. 57–63, Nov. 2021.
[67] SLICES. *Scientific Largescale Infrastructure for Computing/Communication Experimental Studies*. Accessed: Jul. 9, 2023. [Online]. Available: https://www.slices-ri.eu/

LÁSZLÓ TOKA (Member, IEEE) received the Ph.D. degree from Telecom ParisTech, in 2011. He was with Ericsson Research, from 2011 to 2014, then he joined the academia with research focus on software-defined networking, cloud computing, and artificial intelligence. He is currently an Associate Professor with the Budapest University of Technology and Economics, the Vice-Head of the HSN Laboratory, and a member of the ELKH-BME Network Cloud Applications Research Group.

MÁRK KONRÁD (Graduate Student Member, IEEE) received the B.Sc. degree in computer engineering, with a focus in distributed cloud-native storage systems from BME, where he is currently pursuing the master's degree. His current academic research interests include federated learning, edge computing, and 6G technologies.

ISTVÁN PELLE is currently pursuing the Ph.D. degree with the Budapest University of Technology and Economics (BME). He is also a member of the MTA-BME Network Softwarization and ELKH-BME Cloud Applications Research Groups, where he works on cloud and edge computing, with a special focus on services leveraging the serverless concept and Amazon Web Services.

BALÁZS SONKOLY received the M.Sc. and Ph.D. degrees in computer science from the Budapest University of Technology and Economics (BME), in 2002 and 2010, respectively. He is an Associate Professor with BME and the Head of the MTA-BME Network Softwarization Research Group. He has participated in several EU projects (FP7 OpenLab, FP7 UNIFY, and H2020 5G Exchange) and national projects. His current research activity focuses on cloud computing, edge computing, fog computing, NFV, SDN, and 5G. He was the Demo Co-Chair of ACM SIGCOMM 2018, EWSDN'15,'14, and IEEE HPSR'15.

MARCELL SZABÓ received the B.Sc. degree in computer engineering from the Budapest University of Technology and Economics, in 2022, where he is currently pursuing the master's degree. He is also a Student Researcher with Ericsson Hungary. He participates in both academic and industrial research projects, which share machine learning and edge computing as their core field.

BHAVISHYA SHARMA received the B.Tech. degree in computer science from the University of Delhi, India, in 2017. He is currently pursuing the master's degree with the Networked Wireless Systems (NeWS) Laboratory, Indian Institute of Technology Hyderabad (IITH), India. He is also a Research Assistant with NeWS Laboratory, IITH. His research interests include 5G and beyond mobile networks, cloud-RAN, vehicle-to-everything (V2X), network function virtualization, and cloud native technologies.

SHASHWAT KUMAR received the B.Tech. degree in information technology from Uttar Pradesh Technical University, India, in 2011, the M.Tech. degree in computer science and engineering from the Sardar Vallabhbhai National Institute of Technology, Surat, India, in 2013, and the Ph.D. degree in computer science and engineering from the Indian Institute of Technology Hyderabad (IITH), India, in 2022. His current research interests include multi-access edge computing (MEC), mobile networks (5G and beyond 5G), multimedia services optimization, deep learning, and reinforcement learning.

MADHURI ANNAVAZZALA received the B.Tech. degree in computer science and engineering from the Malla Reddy Engineering College for Women. She is currently pursuing the master's degree with the Networked Wireless Systems (NeWS) Laboratory, Indian Institute of Technology Hyderabad (IITH), India. She is also a Research Assistant with the NeWS Laboratory, IITH. Her research interests include wireless networks (5G and beyond), mobile edge computing (MEC), satellite networks, vehicle-to-everything (V2X), and HD maps for autonomous driving. She received "The Young Engineer's Award" from IITH.

SREE TEJA DEEKSHITULA received the B.Tech. degree from the Vardhaman College of Engineering (VCEH), India, in 2020. She is currently pursuing the master's degree with the Networked Wireless Systems (NeWS) Laboratory, Indian Institute of Technology Hyderabad (IITH), India. Her current research interests include vehicle-to-everything (V2X) communication, autonomous vehicles, and caching and resource scaling in cloud native environments.

A. ANTONY FRANKLIN (Senior Member, IEEE) received the B.E. degree in electronics and communication engineering from Madurai Kamaraj University, India, in 2000, the M.E. degree in computer science and engineering from Anna University, India, in 2002, and the Ph.D. degree in computer science and engineering from the Indian Institute of Technology Madras, India, in 2010. He is currently an Associate Professor with the Indian Institute of Technology Hyderabad (IITH), India. Before joining IITH, he was a Senior Engineer with the DMC Research and Development Center, Samsung Electronics, South Korea, from 2012 to 2015, where he has involved in the development of 5G networking technologies. He was a Research Engineer with the Electronics and Telecommunications Research Institute (ETRI), South Korea, from 2010 to 2012, where he has involved in cognitive radio technology research. His current research interests include mobile networks (5G and beyond 5G), cloud radio access networks (C-RAN), mobile edge computing (MEC), the Internet of Things (IoT), and SDN/NFV. He is a member of ACM. He was a co-recipient of the Best Paper Award at COMSNETS 2022, the Best Academic Demo Award at COMSNETS 2018, and the second Best Paper Award at IEEE ANTS 2017 Conferences. He served as the TPC Co-Chair for the National Conference on Communications (NCC), in 2018, and COMSNETS 2019 (Poster Session) Conferences; and the General Co-Chair for the workshop "5G From Theory to Practice" in 5G World Forum, in 2020.