Mobility Digital Twin with Connected Vehicles and Cloud Computing

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

18-10-2021 / 21-10-2021

CITATION

Wang, Ziran (2021): Mobility Digital Twin with Connected Vehicles and Cloud Computing. TechRxiv. Preprint.
<https://doi.org/10.36227/techrxiv.16828759.v1>

DOI

[10.36227/techrxiv.16828759.v1](https://doi.org/10.36227/techrxiv.16828759.v1)

Mobility Digital Twin with Connected Vehicles and Cloud Computing

Ziran Wang, *Member, IEEE*, Rohit Gupta, Kyungtae Han, *Senior Member, IEEE*, Akila Ganlath, Nejib Ammar, and Prashant Tiwari

Abstract—A Digital Twin is a digital replica of a living or non-living physical entity, and this emerging technology has attracted extensive attention from different industries during the past decade. Although a few Digital Twin studies have been conducted in the transportation domain very recently, there is no systematic research with a holistic framework connecting various mobility entities together. In this study, by leveraging both connected vehicle technology and cloud computing, an Mobility Digital Twin (MDT) framework is developed, which consists of three building blocks in the physical space (namely *Human*, *Vehicle*, and *Traffic*), and their associated Digital Twins in the digital space. The cloud architecture is built with Amazon Web Services (AWS) to accommodate the proposed MDT framework and to implement its digital functionalities of storage, modeling, learning, simulation, and prediction. The effectiveness of the MDT framework is shown through the case studies of three digital building blocks with their key microservices: the *Human Digital Twin* with user management and driver type classification, the *Vehicle Digital Twin* with cloud-based Advanced Driver-Assistance Systems (ADAS), and the *Traffic Digital Twin* with traffic flow monitoring and variable speed limit.

Index Terms—Digital Twin, connected vehicles, cloud computing, intelligent transportation systems, Amazon Web Services

I. INTRODUCTION

THE recent development of the Internet of Things (IoT) has been facilitating all kinds of cutting-edge technologies, where their application scenarios are rooted both in the user level (namely individual consumer or private company), and the system level (namely commercial or industrial sector). From the user's perspective, the introduction of the IoT will play a leading role in scenarios like assisted living, e-health, and enhanced learning. From the system's perspective, the most apparent scenarios will be industrial manufacturing, logistics, business/process management, and intelligent transportation of people and goods [1].

The Digital Twin, as an emerging representation of the IoT or Cyber-Physical Systems (CPS), has attracted increasing attention over the past decade [2]. It was ranked as one of the top 10 strategic technology trends for 2019 according to Gartner [3], among other technologies such as autonomous things (e.g., autonomous vehicles), immersive technologies (e.g., virtual reality and augmented reality), and quantum computing. Based on a market research report, the global Digital Twin market size was valued at USD 3.1 billion in

2020, and is projected to reach USD 48.2 billion by 2026, with a compound annual growth rate of 58% during this forecast period [4]. It was also pointed out in their report that, automotive and transportation industry accounted for the largest market share in the Digital Twin market in 2019.

Although the definitions of the Digital Twin vary in different versions [5], [6], the basic concepts are essentially the same: A Digital Twin is a digital replica of a living or non-living physical entity. Digital Twin technology paves the way to real-time monitoring and synchronization of real-world activities with the virtual counterparts [7]. The Digital Twin concept was first born in the aerospace domain when the National Aeronautics and Space Administration (NASA) adopted that as a key element in its 2010 technology roadmap. Along with its rapid development in different domains during the past decade, including aeronautics and space [5], [8], robotics [9], [10], manufacturing [11], [12], and informatics [13], the Digital Twin also has a huge potential in the transportation domain.

The emergence of connected vehicle technology introduces another platform to implement the Digital Twin. Since the level of connectivity within our vehicles has greatly improved, these equipped vehicles are able to “talk” with other entities, such as with other connected vehicles through vehicle-to-vehicle (V2V) communications, with traffic infrastructures through vehicle-to-infrastructure (V2I) communications, and with cloud servers through vehicle-to-cloud (V2C) communications [14]–[17]. Specifically, V2C communications allow connected vehicles to 1) upload their data to the cloud server, enabling Digital Twins to be built in the digital (cyber) world based on their counterparts in the physical world; and 2) offload their onboard computations to the cloud server, enabling Digital Twins to build models and calculate guidance information through powerful cloud computing, which can then be fed back to connected vehicles.

Very recently, a few Digital Twin studies have been conducted in the transportation domain [18]–[20], but none of them has a holistic framework connecting various mobility entities (i.e., human, vehicle, and traffic) together. In this study, an Mobility Digital Twin (MDT) framework is proposed with connected vehicle technology and cloud computing. The MDT framework is built on top of three different layers: 1) the physical space that has human beings, vehicles, and traffic infrastructures; 2) the digital space that has the digital replicas of aforementioned physical entities; and 3) the communication layer between these two spaces. Given the connectivity nature of this framework, it transforms connected vehicles into Internet of Vehicles (IoV) by leveraging IoT

Z. Wang, R. Gupta, K. Han, A. Ganlath, N. Ammar, H. Muraidharan, and P. Tiwari are with Toyota Motor North America R&D, InfoTech Labs, 465 Bernardo Avenue, Mountain View, CA 94043 (e-mail: ryanwang11@hotmail.com; rohit.gupta@toyota.com; kyungtae.han@toyota.com; akila.ganlath@toyota.com; nejib.ammar@toyota.com; prashant.tiwari@toyota.com).

technologies. The cloud architecture is built with Amazon Web Services (AWS) to accommodate the proposed MDT framework and implement its digital functionalities of storage, modeling, learning, simulation, and prediction. The effectiveness of the MDT framework is shown through the case studies of three digital building blocks with their key microservices: the *Human Digital Twin* with user management and driver type classification, the *Vehicle Digital Twin* with cloud-based Advanced Driver-Assistance Systems (ADAS), and the *Traffic Digital Twin* with traffic flow monitoring and variable speed limit.

Since traditional mobility system frameworks heavily rely on onboard storage and computing, their functionalities are limited by multiple constraints, such as computing power, accessibility to big data, and easiness of deployments and modifications. On the contrary, the proposed MDT framework in this study addresses these constraints by making the following contributions:

- **Powerful:** The MDT framework allows users to rapidly adjust cloud resources to meet fluctuating/unpredictable demands, providing high computing power at certain periods of peak demand.
- **Shareable:** Bulk data generated by an end user is offloaded and stored on the cloud, which can be retrieved and utilized by the same user at a later time frame, or shared with other end users for microservices on demand.
- **Manageable:** The MDT framework allows users to get their microservices up and running faster on the cloud platform, with improved manageability and less maintenance. Over-the-air (OTA) updates are also available to the MDT framework.
- **Extendable:** Arbitrary mobility microservices can be easily implemented to the MDT framework with minimal change on the cloud architecture and data structure.

The remainder of this study is organized as follows: Section II conducts a literature review regarding cloud computing and the Digital Twin in the context of connected vehicles. Section III introduces the framework of this MDT with a detailed explanation of three layers: communication layer (and data workflow), physical space, and digital space. Then, the cloud architecture based on AWS is developed in Section IV, which accommodates the proposed MDT framework. Finally, this study is finished with a brief conclusion in Section VI and a discussion about future challenges in Section VII.

II. LITERATURE REVIEW

A. Transportation Applications with Cloud Computing

The emergence of commercial cloud computing services, such as Amazon Web Services (AWS) [21], Microsoft Azure [22], Google Cloud Platform (GCP) [23], and Alibaba Cloud [24], has facilitated many applications in the domain of vehicular/transportation CPS. Such services always provide a variety of basic abstract technical infrastructure, and building blocks for distributed computing. Taking AWS as an example, which has the largest market share among all competitors in 2020, it comprises over 200 products and services for computing, storage, networking, database, analytics, IoT, and so on [21].

All these features of cloud computing services, together with their advantage of scalability, enable connected vehicles to offload their data and onboard computing demand to the cloud.

Guerrero et al. demonstrated that cloud computing can be integrated with intelligent transportation systems to address issues faced by the transportation sector, such as traffic congestion, roadway safety, and pollutant emissions [25]. Specifically, the concept of vehicular cloud can enhance transportation systems by storing and processing the collected data (including traffic lights, parking meters, camera images, etc.), and creating a historical registry of various data sources [26]. Therefore, the transportation authorities who own these entities can make informed decisions on when to change traffic directions, install new traffic lights, and remodel/repair road segments. However, the detailed cloud architecture design is not covered in these studies, and the vehicular cloud applications are introduced only on the conceptual level without conducting case studies.

During the past decade, various transportation applications have been proposed by leveraging the capability of cloud computing [27], [28]. A navigation-assisted route optimizer was developed by Gerla, where the navigator server collects information from connected vehicles, and then computes the optimal routes by constructing a traffic load map and traffic pattern matrix, estimating road segment loads and delays [29]. A bus smart sensor prototype was designed and implemented by Herrera-Quintero et al. using the serverless and microservice cloud architecture, where GCP Firebase was used for storage and AWS Lambda was used for computation [30]. A vehicular pollutant emission detection system was developed by Bhatnagar et al., where AWS IoT and Amazon DynamoDB were integrated to send notifications to the vehicle driver if the emission sensor detects a gas leakage [31]. A vehicle-based traffic surveillance application was developed by Deng et al., where the AWS-based serverless cloud architecture was proved to be feasible for real-time transportation applications through a field implementation [32]. However, aforementioned studies focus more on individual transportation application that provides solutions within a very limited domain, while none of them designs a holistic framework that connects various mobility entities, benefiting human, vehicle, and traffic at the same time.

B. Digital Twin Framework for Connected Vehicles

The Digital Twin concept has been loosely defined and adopted in the transportation domain since its emergence, partly due to its similarity and connection with other technologies. However, many previous efforts related to the IoT and CPS in the automotive industry envision the development of the Digital Twin, since the majority of those proposed methodologies and/or algorithms were developed on multi-layer system frameworks with physical entities (i.e., vehicles) and their digital replicas (simulation models/environments).

Alam and Saddik developed a Digital Twin framework reference model for the cloud-based CPS, where a telematics-based driving assistance application was proposed for the vehicular CPS consisting of three parts: 1) computation, 2) control, and 3) sensors and services fusion [18]. Kumar et

Fig. 1. Illustration of the Mobility Digital Twin (MDT) framework for connected vehicles, which consists of a physical space, a digital space, and a communication layer between two spaces.

al. proposed a Digital Twin-centric approach with machine learning, edge computing, 5G communication, and data lake, aiming for driver intention prediction and traffic congestion avoidance [19]. Chen et al. proposed a “Digital Behavior Twin” framework in which behavioral models of drivers are shared among connected vehicles to predict future actions of neighboring vehicles and hence improve driving safety [20]. This idea was extended to two subsequent patent applications by the same authors [33], [34]. However, the term of “Digital Twin” is used more like a simple concept in these studies, where their presented technologies are agnostic to any IoT technologies besides the Digital Twin. They do not systematically integrate the Digital Twin in any cloud architecture design, nor do they have clear data structures or workflows.

Related literature in the fields of “parallel driving” or “parallel transportation” has direct implications to the Digital Twin framework of our study. In 2010, Wang brought up the parallel transportation concept for the first time, where he defined parallel control and management of transportation as “a data-driven approach for modeling, analysis, and decision-making that considers both the engineering and social complexity in its processes” [35]. Many subsequent works were conducted in this research domain afterward, including the parallel driving framework proposed by Wang et al. [36]. In this cloud-based cyber-physical-social system framework, the physical world, mental world, and artificial world are modeled as three parallel levels, considering interactions among connected vehicles,

human drivers, and information. However, many applications shown in these studies do not come up with a cloud-based framework, and in such cases their capabilities of storage, learning, and prediction are well limited by vehicle onboard resources.

More relevant studies have been conducted very recently by the authors in the context of the Digital Twin for connected vehicles. Wang et al. proposed a Digital Twin paradigm for an advanced driver-assistance system (ADAS) of connected vehicles [37]. In this paradigm, onboard devices on connected vehicles collect and upload data to the cloud server through cellular-based V2C communication, where the cloud server can create digital replicas of entities in the real world (i.e., roads, vehicles, and drivers) based on the received data. All proposed models and algorithms are applied to these digital copies with cloud computing, where their results are propagated back to the real connected vehicles through V2C communication for ADAS, assisting the decision making of drivers in real time. A subsequent field implementation using this Digital Twin paradigm was conducted by Liao et al., where three human-driven passenger vehicles performed ramp merging cooperatively, showing the benefits of safety and environmental sustainability compared to the traditional ramp merging scenario [38]. Visualization of the Digital Twin information from the cloud remains a challenging issue, where Liu et al. developed a data-fusion methodology to overlay the Digital Twin information for the driver’s field of view with the

help of cameras (RGB and depth) images, assisting the driver to make lane change prediction of the neighboring vehicles [39]. However, none of these recent studies, which are from the authors of this study as well, designs a holistic system framework that connects mobility entities (i.e., human, vehicle, and traffic) together, and neither do they develop any cloud architecture with detailed data structure and workflow.

It needs to be noted that, many studies consider the Digital Twin simply as a high-fidelity modeling and simulation environment of real-world entities. Although this statement is partially correct, our understanding of the Digital Twin covers wider than merely modeling and simulation, namely sampling and actuation in the physical space, and storage, modeling, learning, simulation, and prediction in the digital space. Related literature with the limited definition of the Digital Twin is not reviewed in this study.

III. MOBILITY DIGITAL TWIN FRAMEWORK

The MDT framework proposed in this study, as shown in Fig. 1, consists of three layers: 1) The lower layer, highlighted in yellow, stands for the physical space where human beings, vehicles, and traffic infrastructures reside; 2) The upper layer, highlighted in blue, represents the digital space where the digital replicas of those physical entities are located at; 3) Between these two layers, the communication layer (in grey) plays a crucial role in this framework to allow real-time and non-real-time data streaming for both upstream and downstream.

Three entities are considered in this MDT framework: *Human*, *Vehicle*, and *Traffic*. Given the existence of the communication layer, each of the entity can be connected to the digital space (e.g., Internet) and exchanges data with each other. Therefore, this MDT framework is a good representation of the IoT, and it allows connected vehicles to act as IoVs with IoT technologies. In this section, we provide a deep dive into this MDT framework regarding all three aforementioned layers, introducing their building blocks with respect to their definitions and functionalities.

A. Communication Layer and Data Workflow

The communication layer of this MDT framework sits between the physical space and digital space, and it provides seamless connections between these two spaces. This MDT framework's end-to-end process starts from sampling data in the physical space, where all or part of the data is transmitted upstream to the digital space via the communication layer. Those data will go through one or multiple processes in the digital space internally, including storage, modeling, simulation, learning, and prediction, and the resulting data is transmitted downstream to the physical space via the communication layer. Those data, upon receiving, is applied by the actuators of the physical space to fulfill the end-to-end process.

Since cloud computing is leveraged in this MDT framework, the digital space of the framework is deployed fully or partially on the commercial and/or private cloud. Therefore, the communication module needs to provide access to the cloud for the physical space, which is either direct access or indirect access (via edges). The MDT framework does not necessarily require

any specific wireless communication technology to be served as the communication layer, as long as it can be applied to transmit data between the physical space and the digital space.

B. Physical Space

If we consider this MDT framework as an end-to-end framework, then the physical space of this framework is in charge of both ends of this framework, namely, sampling and actuation. We assume no (or only minimal) computing work needs to be conducted in the physical space, since all (or majority) of that is offloaded to the digital space through communication.

For sampling, sensors in the physical space detect the dynamic status, operating process, or event occurrences, and then aggregate these measurements under various resolutions for their transmission to the digital space. On the other hand, once the processed results are received from the digital space, actuation can be made by physical entities to fulfill this end-to-end framework. Generally, the physical space is defined on a world coordinate, which may contain all the transportation-related physical entities, and can be classified into three building blocks: *Human*, *Vehicle*, and *Traffic*.

1) *Human*: In this framework, all human beings involved in the transportation system are considered, which include not only drivers but also passengers, pedestrians, cyclists, etc. The sampling process can be accomplished by the human-machine interface as an active manner, or by the in-cabin status sensing (e.g., camera, seat sensor, etc.), human wellness monitor (e.g., smartwatch, electrocardiogram, etc.), and other perception sensors as a passive manner. The preferences of a human's behavior can also be set actively (e.g., a driver manually sets the preferred cruise control speed), or be measured passively (e.g., a pedestrian's preferred trajectory of crossing a crosswalk is recorded by the vehicle/intersection camera), where both of them are considered as the sampling process.

The actuation process of the *Human* block in this MDT framework is mainly conducted by drivers. In the foreseeable future, our transportation system will remain in a mixed autonomy traffic environment, where only part of all vehicles will be fully autonomous vehicles (with SAE level-5 automation), but the majority are still driven by human drivers (with no degree or a certain degree of automation). Therefore, if drivers can be provided with additional information from the digital space of this MDT framework, such as adjacent vehicle's lane-change possibility or upcoming signal timing, their actuation will be more accurate and in turn benefit other entities in the transportation system.

2) *Vehicle*: *Vehicle* is the core of this MDT framework, as it is the host of drivers and passengers, and also the fundamental component of traffic. As can be seen from Fig. 1, all modules in the physical space, not only the ones in the *Vehicle* block itself but also those in the *Human* block and the *Traffic* block, are serving for vehicle-related activities.

Specifically for the *Vehicle* block, the localization module (i.e., GNSS), the perception sensors (i.e., ultrasonic, camera, radar, and/or LiDAR), together with vehicle CAN BUS are in charge of the sampling process. Related data, such as

positions, speeds, and accelerations of the ego vehicle and its surrounding vehicles can be sampled from these physical components, and then be propagated to the digital space through communication.

The actuation process of the *Vehicle* block in this MDT framework is conducted by the vehicle steering system, accelerator, and brake. These physical components are able to actuate any lateral or longitudinal control command received from the digital space, and therefore allow the vehicle to achieve its desired motion.

3) *Traffic*: Many existing intelligent vehicle platforms and applications, such as ADAS or autonomous driving systems (ADS), only focus on their performances on the ego vehicle without considering their interactions with the large-scale traffic network. However, as can be seen from Fig. 1, *Traffic* is indeed a crucial building block of our MDT framework for connected vehicles. The beneficiaries of this MDT framework include not only connected vehicles and their occupants, but also the whole traffic network on a wider scope.

Particularly, the *Traffic* block in the physical space includes various traffic infrastructures, such as traffic signals, roadside units, camera/radar/loop detectors, and electronic traffic signs. These physical components are able to either generate data (e.g., signal phase and timing) by themselves, or measure data (e.g., traffic count and traffic flow) generated by other traffic entities. Such data is sampled and sent to the digital space through communication, benefiting other building blocks of this MDT framework.

On the other hand, guidance or adjustment received from the digital space can also be actuated by the *Traffic* block to improve the safety and efficiency of the large-scale traffic network. For example, the signal phase and timing of traffic lights can be adjusted to better serve different traffic flows under different situations. Guidance or warning information can be broadcast to connected vehicles via roadside units, and to all traffic entities via electronic traffic signs.

C. Digital Space

The aforementioned physical space of this MDT framework handles both ends of this end-to-end framework (i.e., sampling and actuation). On the other hand, the digital space is in charge of the processes between both ends: storage, modeling, learning, simulation, and prediction.

One of the biggest strengths of this MDT framework over traditional mobility system frameworks is the data lake, which is a centralized repository that allows structured or unstructured data at any scale to be stored. Traditionally, mobility data measured by a physical entity is only saved in its onboard data storage due to the lack of communication capability. Such data is only used for the physical entity itself without being shared with other entities, and will be wiped out once the maximum size limit of the onboard data storage is met. However, with the proposed MDT framework, mobility data measured by the *Human*, *Vehicle*, and *Traffic* blocks in the physical space can be transmitted to the digital space through the communication layer, and stored in the data lakes of associated Digital Twins for future use. Such data can be used for the microservices not

only in the original mobility block, but also in other blocks (e.g., traffic signal data measured by the *Traffic* block can be used for both the “real-time monitoring” microservice in the *Traffic Digital Twin* and the “cooperative control” microservice in the *Vehicle Digital Twin*).

Note there exists a misunderstanding about the Digital Twin in the research community, where some simply consider Digital Twin technology as a modeling and simulation technology. In our MDT framework, modeling and simulation are part of the digital processes that are enhanced by the data lake and data sharing in the digital space, where co-simulation platforms can be built to synchronize data from multiple simulators (such as the Unity-SUMO integrated platform [40]). However, as shown in Fig. 1, our MDT framework is more than just modeling and simulation, where other digital processes (i.e., storage, learning, and prediction) play equivalent roles in the digital space. All of aforementioned digital processes can be applied to mobility microservices, and they are realized in a more powerful, shareable, manageable, and extendable manner by leveraging the cloud architecture and cloud computing.

1) *Human Digital Twin*: *Human Digital Twins* are digital replicas of real humans in the physical space. This building block in the digital space has a human data lake that stores all data sampled from the *Human* block in the physical space, where different humans have their personal databases to be differentiated from others. With real-time data sampling and historical data storage, the *Human Digital Twin* is able to classify drivers into specific driver types by machine learning algorithms like k-nearest neighbors (KNN), and to provide guidance in a customized or personalized manner [41]. Taking advantage of the data coming from the *Vehicle* block, the *Human Digital Twin* can also predict future behaviors of drivers (e.g., lane-change intention [42]) and detect their anomalies [43]. The results of the aforementioned microservices can be applied to third parties such as insurance companies, where they can further build a microservice to set the insurance pricing for different drivers based on their driving behaviors [44].

2) *Vehicle Digital Twin*: *Vehicle Digital Twins* are the digital replicas of real vehicles in the physical space. Once the sampled data is received from a connected vehicle in the physical space, it can be saved in this particular vehicle’s data lake with a unique identification number. Those data in the *Vehicle Digital Twin* about the ego vehicle (e.g., position, speed, and acceleration) and its surrounding environment (perceived by perception sensors) can also be shared with the *Human Digital Twin*, the *Traffic Digital Twin*, or other connected vehicles’ *Vehicle Digital Twins* for various microservices.

With massive data storage and data sharing in the digital space, multiple vehicle-related microservices can be enabled, such as the ones requiring cooperation among multiple connected vehicles: cooperative localization [45], cooperative perception [46], cooperative planning [47], and cooperative control [14]. Additionally, microservices that need time-series data can also be benefited from this MDT framework, where one typical example is predictive maintenance: Based on modeling and simulation of the time-series vehicle data that

Fig. 2. Cloud architecture of the proposed Mobility Digital Twin framework with AWS.

is sampled from the *Vehicle* block in the physical space and stored in the *Vehicle Digital Twin*, the learning process can be conducted in the digital space and predictions can be made regarding potential failures of vehicle components at a future time [48]. Such prediction results can be used by the vehicle owner or manufacture to schedule onsite maintenance before the components break down.

3) *Traffic Digital Twin*: *Traffic Digital Twins* are the digital replicas of traffic infrastructures, which receive data from the *Traffic* block in the physical space. Such sampled data, like signal phase and timing, traffic count, and traffic flow, can be stored in the traffic data lake for future reference. It can also be used for multiple traffic microservices in real time, such as monitoring the traffic condition [19], variable speed limit [49], routing and navigation [50], ridesharing planning [51], and parking management [52].

Similar to the *Human Digital Twin* and *Vehicle Digital Twin*, the *Traffic Digital Twin* can be enhanced by the communication among these Digital Twin blocks. For example, the microservice of routing and navigation can be carried out solely by the real-time traffic flow data sampled from camera/radar/loop detectors in the real world. However, they can be further enhanced if behavior preferences are set by the *Human* block and predictions are made by the *Human Digital Twin* (e.g., a driver/passenger always goes to grocery stores when his/her commute route is highly congested). Additionally, if the *Vehicle* block detects the fuel/battery level is low and sends that to the *Vehicle Digital Twin*, it can also assist the routing and navigation microservice to find a gas/charging station near a user-preferred grocery store along the original route.

IV. CLOUD ARCHITECTURE WITH AWS

In this section, we build the cloud architecture with AWS that can accommodate our proposed MDT framework for con-

nected vehicles. Our purpose is to build a data-driven platform for both real-time and bulk-batch ingestion, processing, and analytics. As shown in Fig. 2, the cloud architecture can be divided into three parts: AWS, Amazon Virtual Private Cloud (VPC) [53] located inside of AWS, and the external space located outside of AWS.

In Amazon VPC, there are five different modules: Distributed Real-Time Processing, Data Stores, Analytics Workbench, Rule Engine, and AI/ML Framework & Digital Twin Microservices. Specifically, in Distributed Real-time Processing, we leverage existing tools like Amazon EKS [54], Apache Kafka [55], and Apache Storm [56] to provide real-time processing and analytics. The Analytics Workbench is the workhorse for big data analytics. It consists of OpenTSDB, which is a distributed, scalable, time-series database built on top of Hadoop and HBase [57]. This supports a writing rate of up to millions of entries per second, supports data storage with millisecond-level precision, and preserves data permanently without sacrificing precision. In addition, Apache Spark [58], a distributed processing system, is used to conduct predictive analytics using Amazon EMR clusters [59].

Rule Engine service evaluates the rules configured for entities (e.g., humans and vehicles) on the data received from the Kafka queue, and redirects it to AI/ML Framework & Digital Twin Microservices based on the rule validation result. AI/ML Framework & Digital Twin Microservices are the core of this cloud architecture, where end users are able to implement customized algorithms and applications with various objectives. This module processes time-series data sent from the physical space using statistical techniques, and sends guidance back to the entities in the physical space. The data workflow is triggered via Apache Airflow, an open-source workflow management platform.

Data Stores of our cloud architecture are made up with:

1) Amazon S3, a scalable storage infrastructure to build our Digital Twin data lake [60]; 2) Amazon DocumentDB (with MongoDB compatibility), a database service that is purpose-built for JSON data to execute flexible, low latency queries to obtain a near real-time record of events in parallel on a massive scale [61]; and 3) Redis, an open-source, highly replicated, non-relational kind of database and caching server [62].

Outside of Amazon VPC but inside of AWS sits AWS IoT Core [63], which enables the connection between IoT devices (such as mobile apps, simulators, real vehicles, and RC vehicles in this study) and AWS cloud without the need to provision or manage servers. It supports various devices and messages, and can process/route those messages to AWS endpoints/devices reliably and securely. A Bulk Data Ingestion module is also developed in this cloud architecture, enabling the ingestion of terabytes of data in batch mode into our data lake. Some scenarios where this module can be triggered are 1) end of vehicle trip bulk data ingestion; 2) periodic bulk data ingestion; 3) event-triggered data ingestion; 4) in-vehicle data logging.

OpenID Connect is a simple identity layer on top of the OAuth 2.0 authorization protocol, which is adopted in this cloud architecture to verify the identity of end users based on the authentication performed by an authorization server, as well as to obtain the basic profile information about end users [64]. Amazon API (namely application programming interface) Gateway, as an AWS managed service, is adopted to create, publish, maintain, monitor, and secure APIs at any scale [65]. The API Gateway acts as the “front door” for applications to access data or functionalities from our back-end services.

Outside of AWS, external data sources can be leveraged to enrich the functionalities of cloud microservices. For example, traffic data (TomTom [66]), map data (OpenStreetMap [67]), and weather data (OpenWeather [68]) are integrated into AWS API Gateway via HTTP. With such data, more microservices can be deployed in the *Traffic Digital Twin*, and hence provide better guidance towards humans and vehicles in the physical space. Additionally, a web portal is designed to visualize the digital processes on the cloud, and enable end users to create and modify microservices.

The most important data sources are shown on the right side of Fig. 2, which stand for the *Human*, *Vehicle* and *Traffic* building blocks in the physical space of the MDT framework. Mobile apps are designed for both Android and iOS, where end users’ position and speed data (measured by GPS and gyroscope) can be uploaded to AWS IoT Core via MQTT. Additionally, a customized edge gateway is developed to allow external simulators (such as SUMO [69] and Unity [70]), and real vehicles and RC vehicles (with ROS2 embedded) to transmit messages with AWS IoT Core via MQTT. To the current stage of our study, these are the ways that end users can generate data in the physical space, and they will be further introduced in the next section.

Fig. 3. Real-world demonstration in Mountain View, CA of the cloud-based driver scoring system, used for user management and driver type classification.

Fig. 4. User management page on the web portal of the Mobility Digital Twin, showing user’s driver risk class (among reckless, intermediate, and competent), associated vehicle (currently and historically), and past trips with clickable links (in light blue fonts) for more details.

V. CASE STUDIES OF THE MOBILITY DIGITAL TWIN

A. Human Digital Twin: User Management and Driver Type Classification

In this subsection, we first conduct a case study on the *Human Digital Twin*. A web portal is built to enable various data management and visualization functionalities, where related data sampled in the physical space is transmitted and visualized on the web portal through HTTPS, as illustrated in Fig. 2.

To begin with, each human user needs to register a unique account of the MDT through the web portal, and also to specify the group(s) he/she belongs to: “viewer”, “supervisor”, and/or “admin”. Particularly, the user account that belongs to the “admin” group is granted full access to the human data lake, which can register/delete any account in the digital space, alter the group(s) any account belongs to, and subscribe/unsubscribe microservices for any account. “Supervisor” account can only access its own human data lake with modification right, while “viewer” account has no modification right at all.

Human-vehicle association is also available through the web portal, which enables the human user to associate his/her vehicle in the vehicle data lake, building the connection between the *Human Digital Twin* and the *Vehicle Digital Twin*. Therefore, once the user account is registered, all the data generated by this human user in the physical space will be stored both in the user data lake (under the particular user ID) and in the vehicle data lake (under the particular vehicle ID).

Besides the aforementioned details regarding user management, the built-in microservices of the *Human Digital Twin* can also be applied. One example is shown as Fig. 3, where different driving scores of a driver (i.e., overall score, eco score, safe score, and comfort score) are calculated by the open-source MOVESTAR model [71] running on the cloud in real time. These driving scores can be further compared with the historical data, which is generated by other drivers and stored in the human data lake to classify this driver into a certain type. An example classification result is visualized on the web portal (currently shown as “Competent” in Fig. 4), which can be used for other microservices such as behavior prediction, personalized guidance, and insurance pricing. The historical classification results of this driver can also be retrieved by clicking the detailed trip list shown in Fig. 4, which further validates the power of the *Human Digital Twin* in terms of storage.

B. Vehicle Digital Twin: Cloud-Based Advanced Driver-Assistance Systems

In this subsection, we conduct a case study on the *Vehicle Digital Twin* through a cloud-based ADAS, where cloud computing is leveraged to provide personalized guidance and control commands towards connected vehicles. As shown in Fig. 5, a human-in-the-loop simulation (built with AWS, the Unity game engine, and the Logitech G29 Driving Force) is conducted to simulate the data sampling process from a connected vehicle in the physical space [72]:

When a human driver manually controls the vehicle, its data is sampled from the physical components (e.g., CAN BUS, radar, camera) and uploaded to the data lake of the *Vehicle Digital Twin*. This process is shown at the lower-left corner “Unity-AWS Uplink Message” of Fig. 5.

The data stored in the data lake (potentially from all historical trips) of this vehicle is inputted to microservices of the *Vehicle Digital Twin* (e.g., cooperative planning, cooperative control, etc.). Machine learning-based algorithms are implemented to learn the performance and preference of each vehicle and/or driver, where the algorithm outputs may include prediction/guidance of their current status and future behaviors.

Such algorithm outputs are downloaded from the *Vehicle Digital Twin* to the vehicle, as shown at the lower-right corner of Fig. 5 “AWS-Unity Downlink Message”. By leveraging computer vision technologies, this information is overlaid on top of each vehicle through an augmented reality (AR) head-up display (HUD) design, assisting the decision making of other drivers [73]. As shown in Fig. 6, from this driver’s field of view, the following information of the surrounding vehicles

Fig. 5. Human-in-the-loop simulation of the cloud-based ADAS based on AWS, the Unity game engine, and the Logitech G29 Driving Force.

Fig. 6. Prediction and guidance information received from the *Vehicle Digital Twin* on the cloud is visualized through an augmented reality (AR) head-up display (HUD), which may include: driving proficiency score and its trend, potential action (e.g., hard braking or lane change) and its possibility, as well as driving mood score.

and their drivers can be known (from top to bottom of the overlaid information): driving proficiency score and its trend, potential action (e.g., hard braking or lane change) and its possibility, as well as driving mood score.

Compared to a traditional ADAS that relies on pure onboard sensing and processing of the ego vehicle, the key advantages that MDT brings to this cloud-based ADAS are:

- Heavy computations, such as training a machine learning algorithm based on the sampled data, can be offloaded to the cloud to utilize more computing power and hence save time.
- Additional data sources in the physical space, such as surrounding vehicles and downstream traffic, can be utilized to enhance the functionalities of ADAS.
- The cloud-based ADAS can be easily migrated through the cloud and hence increase accessibility, which can be available on various vehicles for the same driver, or for various drivers on the same vehicle.
- Updates of ADAS algorithms and applications can be conducted much quicker and easier through OTA updates.

C. Traffic Digital Twin: Traffic Flow Monitoring and Variable Speed Limit

This subsection conducts a case study on the *Traffic Digital Twin* with the microservices of traffic flow monitoring and variable speed limit. The mobile app we developed for MDT,

Fig. 7. (a) Physical space: the iOS mobile app for the Mobility Digital Twin (visualized at the initialization step) that allows end users to upload data to AWS; (b) Digital space: traffic flow monitoring page of the web portal that visualizes the data received from end users; (c) Digital space: variable speed limit based on geo-fences.

together with the microscopic traffic simulator SUMO, is adopted in this case study to generate traffic flow and hence represent connected vehicles traveling in the physical world.

As shown in Fig. 7(a), we develop an iOS mobile app (as well as an Android mobile app that is not shown here) for MDT that allows end users to upload their data from the physical space to the digital space, which may include latitude, longitude, and speed. At the initialization step of the app, the user is asked to associate with a vehicle in the digital space, and also set the data push rate. Once “START” button is pressed, the app will get the data from the mobile phone itself, and push that to the *Vehicle Digital Twin* through AWS IoT Core. Similarly, when SUMO simulation gets started, each vehicle in the simulation publishes its latitude, longitude, and speed to the *Vehicle Digital Twin* through the edge gateway.

When the traffic flow monitoring microservice is turned on, it gets the data from all *Vehicle Digital Twins* and aggregates it in the *Traffic Digital Twin*. A demonstration of this microservice running in motion is shown in the snapshot Fig. 7(b). The count of all vehicles running on the specific link is calculated, and then gets further divided by the link length and number of lanes to get the traffic density value. This value is compared with the pre-defined value ranges for heavy congestion (red), moderate congestion (orange), or no congestion (green) to determine the traffic density on that link and gets represented in the corresponding color on the map. Additionally, the different colors of vehicles visualized on the map indicate their current speeds, where the ones traveling equal or below the average link speed are shown in blue color, and the ones above the average link speed are in white.

In order to better manage traffic flows in the physical space, the variable speed limit can be applied to connected vehicles through the *Traffic Digital Twin*. As can be seen from Fig. 7(c), users can easily draw a geo-fence by specifying its name, radius, latitude, and longitude of the center location. A hexagon will then be visualized on the map, and users are prompted to choose the rules that are applied to vehicles, together with their states (“entry” or “exit”) and associated

Fig. 8. (a) Mass-produced vehicle that is used for the end-to-end latency test; (b) A 10-minute trip conducted on Highway 237 in California to collect CAN BUS data from the vehicle.

TABLE I
LATENCY OF A NON-REAL-TIME APPLICATION THAT PROCESSES 10-MIN CAN BUS DATA AND QUERIES THE HISTORICAL DATABASE

Data Frequency	Uplink Data Total Size	Uplink Latency	Cloud Computing+ Downlink Latency	Overall Latency
0.5 Hz	12 MB	1.4 s	16.1 s	17.5 s
1 Hz	24 MB	2.1 s	16.1 s	18.2 s
5 Hz	120 MB	10.6 s	16.1 s	26.7 s
10 Hz	240 MB	21.1 s	16.1 s	33.2 s

values. Taking the variable speed limit microservice as an example, users can set the rule state “entry” with a value “40” and another rule state “exit” with a value of “60”. This setting indicates that all connected vehicles which subscribe to this microservice will receive a speed recommendation/limit of 40 mph when entering this geo-fence, and 60 mph while exiting.

D. End-to-End Latency Testing with Human/Vehicle/Traffic Digital Twins

In this subsection, we conduct an end-to-end testing of the proposed MDT framework, where the *Human Digital Twin*, *Vehicle Digital Twin*, and *Traffic Digital Twin* all participate in this process. The purpose of this test is to gauge whether the proposed MDT framework can be integrated into mass-produced vehicles for real-world mobility applications. It should be noted that, this testing does not focus on safety-critical or time-critical application, but acts more like a stress testing to verify the stability and reliability of the framework while executing complex computing processes on the cloud.

This end-to-end testing utilizes the CAN BUS data generated from a Lexus LS test vehicle shown as Fig. 8(a), which includes 15 data fields such as position, speed, acceleration, and so on. Time-series data of a 10-minute trip (i.e., Fig. 8(b)) gets sent to AWS through the aforementioned Bulk Data Ingestion module after the trip is finished. On the cloud, multiple processes are conducted for the uplink time-series data, including:

- 1) Store all the raw uplink data into OpenTSDB.
- 2) Invoke statistics service, driver classification service, and weather classification service (with OpenWeather API) for the raw uplink data and store the classification result in MongoDB.
- 3) Filter the raw uplink data to extract the necessary portions, and invoke machine learning algorithms to calculate the outputs.

- 4) Query the MongoDB regarding all historical trips under the same classification result and fetch their learning outputs.
- 5) Compute the downlink output by integrating the learning outputs of this immediate past trip and all historical trips that are under the same classification, and also save the output in MongoDB.

As can be seen from the results in TABLE I, under different sampling frequency of the CAN BUS data, the uplink latency is relatively proportional to the uplink batch size, with a minimum of 1.4 s for 12 MB and a maximum of 21.1 s for 240 MB. However, the majority of the end-to-end latency is made up by the aforementioned cloud computing processes, which roughly takes 16 s under different data frequency. In this end-to-end testing, since the downlink data is pre-defined to follow the same size and format, the major factor of the latency is the number of cloud computing steps instead of the uplink data size.

In a nutshell, the end-to-end testing showcases the capability of the proposed MDT framework to execute complex cloud computing processes based on bulk uplink data within a rational range of latency. In real-time mobility applications like the ones shown in the *Human Digital Twin* as Fig. 3, and in the *Vehicle Digital Twin* as Fig. 5 and Fig. 6, where only a limited number of cloud computing processes are executed based on the uplink data stream, the MDT framework guarantees 80 ms as the medium end-to-end latency to enable safety-critical and time-critical applications.

VI. CONCLUSION

In this study, an MDT framework has been developed with connected vehicles and cloud computing. It has been found from the literature review that, although the Digital Twin concept has been recently studied in the transportation domain, there is no related literature that leverages this technology together with cloud computing to benefit connected vehicles with detailed microservices. In this study, the proposed MDT framework has been demonstrated with details regarding all of its components: *Human*, *Vehicle*, *Traffic*, together with their associated *Human Digital Twin*, *Vehicle Digital Twin*, and *Traffic Digital Twin*.

The public cloud service AWS has been adopted in this study to design the cloud architecture, which accommodates the MDT framework and makes it a reality rather than a concept. To showcase the effectiveness of the MDT framework, case studies have been conducted from all three digital building blocks with their key microservices: the *Human Digital Twin* with user management and driver type classification, the *Vehicle Digital Twin* with cloud-based ADAS, and the *Traffic Digital Twin* with traffic flow monitoring and variable speed limit.

VII. FUTURE CHALLENGES

In the future development and utilization of Digital Twin technology in both academia and industry, together with the involvements of connected vehicle technology and cloud computing, numerous challenges need to be addressed from

the perspectives of both research and engineering. Some of the major challenges are discussed in this section with open questions.

A. Digital Twin Standardization?

Although Digital Twin technology has gained momentum in various domains during the past decade, there is no universal definition of this technology, let alone existing standardization. Currently, the joint technical committee “Internet of Things and Digital Twin” of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), is still developing the standards for the Digital Twin in terms of concepts and terminology [74], as well as use cases [75]. Additionally, the specific standards for the Digital Twin manufacturing framework are also under development by the ISO technical committee “Industrial Data”, with focuses on overview and general principles [76], reference architecture [77], digital representation of manufacturing elements [78], and information exchange [79].

Similar to manufacturing, specific standards need to be developed for the transportation domain, so Digital Twin technology can be fully deployed on the connected vehicles we will be riding in the future. Such standards can be used by different organizations to define APIs for Digital Twin data access, enabling different transportation entities (e.g., human drivers, vehicles, and traffic infrastructures) to securely and reliably store, manage, and retrieve records. Related standards can also help developers design Human Machine Interfaces (HMI) to enable better interactions between physical and digital spaces of the Digital Twin.

However, the standardization of the Digital Twin in the transportation domain can face numerous challenges, since the consensus may be difficult to reach across the public sector (e.g., transportation agencies) and the private sector (e.g., automotive manufacturers, suppliers, and network providers), similar to the everlasting debate between Dedicated Short-Range Communication (DSRC) and Cellular Vehicle-to-Everything (C-V2X) communication for connected vehicles.

B. Public Cloud or Private Cloud?

In this study, a public cloud architecture has been designed and deployed on AWS. However, this does not necessarily mean our proposed MDT framework can only work on AWS instead of other cloud platforms. Other commercial platforms like Microsoft Azure (especially with its “Azure Digital Twins” [80]), GCP and Alibaba Cloud could also be the alternatives to accommodate the MDT framework.

However, a cloud platform must be trustworthy to deploy any of the microservices mentioned in this study (especially for the ones related to the *Human Digital Twin*), as end user information and related data need to be secured from being compromised. The nature of public cloud platforms inevitably gives away the control of resources to some extent, which may introduce cybersecurity and privacy risks for end users.

A private cloud platform, on the other hand, consists of cloud computing resources used exclusively by one business or organization. Since the services and infrastructure are

always maintained on a private network without sharing with others, a private cloud platform can address the security and privacy issues faced by a public cloud platform. Private cloud platforms also make it easier for end users to customize cloud resources to meet specific requirements and implement specific functions, which was proved in our private cloud-based cooperative ramp merging experiment [38].

Although private cloud platforms are more secured and flexible, it has several disadvantages compared to public cloud platforms. In general, private cloud platforms are more expensive, since hardware and software should be dedicated solely to particular organizations that they serve (and hence paid solely by those organizations). In terms of scalability and reliability, private cloud platforms are also outperformed by public cloud platforms, because the public ones provide on-demand resources to meet various organizations' needs, and also provide a vast network of servers to ensure against failure.

Therefore, based on specific requirements and needs of end users, choices can be made between public and private cloud platforms, considering their advantages and disadvantages described above. Additionally, building Digital Twins with a hybrid cloud approach (combining public and private cloud platforms, potentially with edge/fog computing) provides another possibility, which also leads to the discussion in the next subsection.

C. Fully Cloud-Based Approach or Hybrid Approach with Edge Computing?

A fully cloud-based architecture has been designed in this study, where only sampling and actuation processes are conducted in the physical space, but all computation-related processes are conducted on the cloud (i.e., AWS in this case). However, this does not necessarily indicate all digital processes of the proposed MDT framework (shown in Fig. 1) must sit on the cloud.

In fact, it gets difficult sometimes for connected vehicles to have the cloud access, since they continuously move around and may lose internet connection every now and then. Therefore, a hybrid approach with edge/fog computing and cloud computing can meet the requirements of ultra-low latency for running safety-critical microservices at the edge (e.g., road-side units), and of extensive resources for running data-driven microservices on the cloud [28]. Edge computing has already been widely researched by various works in the field of connected vehicles [81], [82], and also deployed in the real world in projects like 5G-MOBIX [83] and initiatives like Automotive Edge Computing Consortium (AECC) [84].

Additionally from the connected vehicle perspective: Instead of dumping all raw data sampled from its physical sensors to the cloud or edge, some onboard post-processing can be done to reduce the data size being uploaded (e.g., retrieving the necessary parameters from raw CAN BUS data), and hence reduce data transmission time and cloud storage cost; Instead of assigning all computing tasks to the cloud or edge, part of the modeling and prediction processes can be done on the vehicle onboard computer (e.g., predicting a target vehicle's behaviors based on a neural network trained

on the cloud), so real-time actuation can be guaranteed with minimum delay. Data and models can be decoupled for transmission between the two spaces of the Digital Twin, where the transmission of data needs more consideration than that of models, since data can be significantly larger in size than model parameters.

To sum up, scalability, reliability, security, cost, and latency are the key factors to consider when decisions are made between the private cloud and the public cloud, and between the fully cloud-based approach and the hybrid approach for Digital Twin frameworks. Along with the rapid development of the Digital Twin, it can be envisioned that more related studies will be conducted in the transportation domain to tackle these challenges in the near future.

ACKNOWLEDGMENT

The authors want to sincerely thank the Atos team for their technical supports in building the AWS cloud framework. The authors want to also thank Shalini Keshavamurthy, Dr. Seyhan Ucar, Haritha Muralidharan, Dr. Takamasa Higuchi, Johnny Wang, Divya Sai Toopran, Akio Orii, and Kazuhisa Shitanaka for their support and guidance for this study.

The contents of this study only reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views of Toyota Motor North America.

REFERENCES

- [1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.
- [2] Y. Guo, X. Hu, B. Hu, J. Cheng, M. Zhou, and R. Y. Kwok, "Mobile cyber physical systems: Current challenges and future networking applications," *IEEE Access*, vol. 6, pp. 12 360–12 368, 2017.
- [3] Gartner. (2018) Gartner top 10 strategic technology trends for 2019. [Online]. Available: <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2019/>
- [4] Market and Markets, "Digital Twin market," Accessed: 2021-01-25. [Online]. Available: <https://www.marketsandmarkets.com/Market-Reports/digital-twin-market-225269522.html#:~:text=%5B177%20Pages%20Report%5D%20The%20global,58%25%20during%20the%20forecast%20period.>
- [5] E. Glaessgen and D. Stargel, "The digital twin paradigm for future NASA and US Air Force vehicles," in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, 2012, p. 1818.
- [6] J. Lee, E. Lapira, B. Bagheri, and H.-A. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," *Manufacturing letters*, vol. 1, no. 1, pp. 38–41, 2013.
- [7] J. Lee, B. Bagheri, and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Manufacturing letters*, vol. 3, pp. 18–23, 2015.
- [8] E. Tuegel, "The airframe digital twin: some challenges to realization," in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*, 2012, p. 1812.
- [9] M. Schluse and J. Rossmann, "From simulation to experimentable digital twins: Simulation-based development and operation of complex technical systems," in *2016 IEEE International Symposium on Systems Engineering (ISSE)*. IEEE, 2016, pp. 1–6.
- [10] G. Grinshpun, T. Cichon, D. Dipika, and J. Rossmann, "From virtual testbeds to real lightweight robots: Development and deployment of control algorithms for soft robots, with particular reference to," in *Proceedings of ISR 2016: 47st International Symposium on Robotics*. VDE, 2016, pp. 1–7.

- [11] R. Rosen, G. Von Wichert, G. Lo, and K. D. Bettenhausen, "About the importance of autonomy and digital twins for the future of manufacturing," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 567–572, 2015.
- [12] G. N. Schroeder, C. Steinmetz, C. E. Pereira, and D. B. Espindola, "Digital twin data modeling with automationml and a communication methodology for data exchange," *IFAC-PapersOnLine*, vol. 49, no. 30, pp. 12–17, 2016.
- [13] A. Canedo, "Industrial IoT lifecycle via digital twins," in *Proceedings of the Eleventh IEEE/ACM/IFIP International Conference on Hardware/Software Codesign and System Synthesis*. ACM, 2016, p. 29.
- [14] Z. Wang, Y. Bian, S. E. Shladover, G. Wu, S. E. Li, and M. J. Barth, "A survey on cooperative longitudinal motion control of multiple connected and automated vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 12, no. 1, pp. 4–24, 2020.
- [15] J. Cheng, G. Yuan, M. Zhou, S. Gao, Z. Huang, and C. Liu, "A connectivity-prediction-based dynamic clustering model for vanet in an urban scene," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8410–8418, 2020.
- [16] Y. Ma, Z. Wang, H. Yang, and L. Yang, "Artificial intelligence applications in the development of autonomous vehicles: a survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 7, no. 2, pp. 315–329, 2020.
- [17] L. Silva, N. Magaia, B. Sousa, A. Kobusińska, A. Casimiro, C. X. Mavroumoustakis, G. Mastorakis, and V. H. C. De Albuquerque, "Computing paradigms in emerging vehicular environments: A review," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 3, pp. 491–511, 2021.
- [18] K. M. Alam and A. El Saddik, "C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems," *IEEE Access*, vol. 5, pp. 2050–2062, 2017.
- [19] S. A. Kumar, R. Madhumathi, P. R. Chelliah, L. Tao, and S. Wang, "A novel digital twin-centric approach for driver intention prediction and traffic congestion avoidance," *Journal of Reliable Intelligent Environments*, vol. 4, no. 4, pp. 199–209, 2018.
- [20] X. Chen, E. Kang, S. Shiraishi, V. M. Preciado, and Z. Jiang, "Digital behavioral twins for safe connected cars," in *Proceedings of the 21th ACM/IEEE International Conference on Model Driven Engineering Languages and Systems*, 2018, pp. 144–153.
- [21] Amazon, "Amazon Web Services (AWS) - Cloud Computing Services," 2021-03-03. [Online]. Available: <https://aws.amazon.com/>
- [22] Microsoft Azure, "Azure Digital Twins," 2021-06-08. [Online]. Available: <https://azure.microsoft.com/en-us/services/digital-twins/>
- [23] Google, "Cloud Computing Services — Google Cloud," 2021-03-03. [Online]. Available: <https://cloud.google.com/>
- [24] Alibaba, "Alibaba Cloud: Reliable & Secure Cloud Solutions to Empower Your Global Business," 2021-03-03. [Online]. Available: <https://www.alibabacloud.com/>
- [25] J. A. Guerrero-Ibanez, S. Zeadally, and J. Contreras-Castillo, "Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and internet of things technologies," *IEEE Wireless Communications*, vol. 22, no. 6, pp. 122–128, 2015.
- [26] S. Olariu, I. Khalil, and M. Abuelela, "Taking vanet to the clouds," *International Journal of Pervasive Computing and Communications*, 2011.
- [27] M. Whaiduzzaman, M. Sookhak, A. Gani, and R. Buyya, "A survey on vehicular cloud computing," *Journal of Network and Computer applications*, vol. 40, pp. 325–344, 2014.
- [28] H. Wang, T. Liu, B. Kim, C.-W. Lin, S. Shiraishi, J. Xie, and Z. Han, "Architectural design alternatives based on cloud/edge/fog computing for connected vehicles," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2349–2377, 2020.
- [29] M. Gerla, "Vehicular cloud computing," in *2012 The 11th annual mediterranean ad hoc networking workshop (Med-Hoc-Net)*. IEEE, 2012, pp. 152–155.
- [30] L. F. Herrera-Quintero, J. C. Vega-Alfonso, K. B. A. Banse, and E. C. Zambrano, "Smart ITS sensor for the transportation planning based on iot approaches using serverless and microservices architecture," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2, pp. 17–27, 2018.
- [31] A. Bhatnagar, V. Sharma, and G. Raj, "IoT based car pollution detection using AWS," in *2018 International Conference on Advances in Computing and Communication Engineering (ICACCE)*. IEEE, 2018, pp. 306–311.
- [32] H. W. Deng, M. Rahman, M. Chowdhury, M. S. Salek, and M. Shue, "Commercial cloud computing for connected vehicle applications in transportation cyberphysical systems: A case study," *IEEE Intelligent Transportation Systems Magazine*, vol. 13, no. 1, pp. 6–19, 2021.
- [33] S. Shiraishi, Z. Jiang, and B. Kim, "Digital twin for vehicle risk evaluation," Dec. 19 2019, US Patent App. 16/007,693.
- [34] Z. Jiang, S. Shiraishi, and B. Kim, "Collision avoidance for a connected vehicle based on a digital behavioral twin," Nov. 24 2020, US Patent 10,843,689.
- [35] F.-Y. Wang, "Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 3, pp. 630–638, 2010.
- [36] F.-Y. Wang, N.-N. Zheng, D. Cao, C. M. Martinez, L. Li, and T. Liu, "Parallel driving in cpss: A unified approach for transport automation and vehicle intelligence," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 577–587, 2017.
- [37] Z. Wang, X. Liao, X. Zhao, K. Han, P. Tiwari, M. J. Barth, and G. Wu, "A Digital Twin paradigm: Vehicle-to-Cloud based advanced driver assistance systems," in *2020 IEEE 91st Vehicular Technology Conference*, May 2020, pp. 1–6.
- [38] X. Liao, Z. Wang, X. Zhao, K. Han, P. Tiwari, M. Barth, and G. Wu, "Cooperative ramp merging design and field implementation: A digital twin approach based on vehicle-to-cloud communication," *IEEE Intelligent Transportation Systems Magazine*, 2021.
- [39] Y. Liu, Z. Wang, K. Han, Z. Shou, P. Tiwari, and J. Hansen, "Vision-cloud data fusion for adas: A lane change prediction case study," *IEEE Transactions on Intelligent Vehicles*, 2021.
- [40] X. Liao, X. Zhao, G. Wu, M. Barth, Z. Wang, K. Han, and P. Tiwari, "A game theory based ramp merging strategy for connected and automated vehicles in the mixed traffic: A unity-sumo integrated platform," *arXiv preprint arXiv:2101.11237*, 2021.
- [41] Z. Wang, X. Liao, C. Wang, D. Oswald, G. Wu, K. Boriboonsomsin, M. Barth, K. Han, B. Kim, and P. Tiwari, "Driver behavior modeling using game engine and real vehicle: A learning-based approach," *IEEE Transactions on Intelligent Vehicles*, pp. 1–1, 2020.
- [42] Z. Shou, Z. Wang, K. Han, Y. Liu, P. Tiwari, and X. Di, "Long-term prediction of lane change maneuver through a multilayer perceptron," in *IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2020.
- [43] M. Zhang, C. Chen, T. Wo, T. Xie, M. Z. A. Bhuiyan, and X. Lin, "Safedrive: online driving anomaly detection from large-scale vehicle data," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 2087–2096, 2017.
- [44] W. Weidner, F. W. Transchel, and R. Weidner, "Telematic driving profile classification in car insurance pricing," *Annals of actuarial science*, vol. 11, no. 2, pp. 213–236, 2017.
- [45] J. Liu, B. Cai, and J. Wang, "Cooperative localization of connected vehicles: Integrating gnss with dsrc using a robust cubature kalman filter," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 2111–2125, 2016.
- [46] S. Aoki, T. Higuchi, and O. Altintas, "Cooperative perception with deep reinforcement learning for connected vehicles," in *2020 IEEE Intelligent Vehicles Symposium (IV)*, 2020, pp. 328–334.
- [47] Z. Wang, K. Han, and P. Tiwari, "Digital Twin-assisted cooperative driving at non-signalized intersections," *IEEE Transactions on Intelligent Vehicles*, 2021.
- [48] T. Masuda, B. Kim, and S. Shiraishi, "Proactive vehicle maintenance scheduling based on digital twin simulations," Aug. 29 2019, US Patent App. 15/908,768.
- [49] M. Papageorgiou, E. Kosmatopoulos, and I. Papamichail, "Effects of variable speed limits on motorway traffic flow," *Transportation Research Record*, vol. 2047, no. 1, pp. 37–48, 2008.
- [50] Y. Pan, N. Wu, T. Qu, P. Li, K. Zhang, and H. Guo, "Digital-twin-driven production logistics synchronization system for vehicle routing problems with pick-up and delivery in industrial park," *International Journal of Computer Integrated Manufacturing*, pp. 1–15, 2020.
- [51] A. K. Tripathy, P. K. Tripathy, A. G. Mohapatra, N. K. Ray, and S. P. Mohanty, "Wedoshare: A ridesharing framework in transportation cyber-physical system for sustainable mobility in smart cities," *IEEE Consumer Electronics Magazine*, vol. 9, no. 4, pp. 41–48, 2020.
- [52] T. Higuchi, S. Ucar, and O. Altintas, "A collaborative approach to finding available parking spots," in *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*. IEEE, 2019, pp. 1–5.
- [53] AWS, "Amazon Virtual Private Cloud," 2021-06-05. [Online]. Available: <https://aws.amazon.com/vpc/?vpc-blogs.sort-by=item.additionalFields.createdDate&vpc-blogs.sort-order=desc>
- [54] —, "Amazon Elastic Kubernetes Service," 2021-06-05. [Online]. Available: <https://aws.amazon.com/eks/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc&eks-blogs.sort-by=item.additionalFields.createdDate&eks-blogs.sort-order=desc>

- [55] Apache, "Apache Kafka," 2021-06-05. [Online]. Available: <https://kafka.apache.org/>
- [56] —, "Apache Storm," 2021-06-05. [Online]. Available: <https://storm.apache.org/>
- [57] OpenTSDB, "The Scalable Time Series Database," 2021-06-05. [Online]. Available: <http://opentsdb.net/>
- [58] Apache, "Apache Spark," 2021-06-05. [Online]. Available: <https://spark.apache.org/>
- [59] AWS, "Amazon EMR," 2021-06-05. [Online]. Available: <https://aws.amazon.com/emr/>
- [60] —, "Amazon S3," 2021-06-05. [Online]. Available: <https://aws.amazon.com/s3/>
- [61] —, "Amazon DocumentDB (with MongoDB compatibility)," 2021-06-05. [Online]. Available: <https://aws.amazon.com/documentdb/>
- [62] Redis, "Redis," 2021-06-05. [Online]. Available: <https://redis.io/>
- [63] Amazon, "AWS IoT Core," 2021-05-31. [Online]. Available: <https://aws.amazon.com/iot-core/>
- [64] OpenID, "Welcome to OpenID Connect," <https://openid.net/connect/>. [Online]. Available: <https://redis.io/>
- [65] AWS, "Amazon API Gateway," 2021-06-05. [Online]. Available: <https://aws.amazon.com/api-gateway/>
- [66] TomTom, "Real-Time Traffic Data," 2021-06-05. [Online]. Available: <https://www.tomtom.com/products/real-time-traffic/>
- [67] OpenStreetMap, "API," 2021-06-01. [Online]. Available: <https://wiki.openstreetmap.org/wiki/API>
- [68] OpenWeather, "Weather API," 2021-06-01. [Online]. Available: <https://openweathermap.org/api>
- [69] Eclipse SUMO, "Simulation of urban mobility," Accessed: 2021-05-03. [Online]. Available: <https://www.eclipse.org/sumo/>
- [70] Unity, "Unity for all." [Online]. Available: <https://unity.com/>
- [71] Z. Wang, G. Wu, and G. Scora, "MOVESTAR: An open-source vehicle fuel and emission model based on USEPA MOVES," *arXiv preprint arXiv:2008.04986*, 2020.
- [72] Z. Wang, K. Han, and P. Tiwari, "Digital Twin Simulation of Connected and Automated Vehicles with the Unity Game Engine," *TechRxiv*, 5 2021. [Online]. Available: https://www.techrxiv.org/articles/preprint/Digital_Twin_Simulation_of_Connected_and_Automated_Vehicles_with_the_Unity_Game_Engine/14582943
- [73] —, "Augmented reality-based advanced driver-assistance system for connected vehicles," in *2020 IEEE International Conference on Systems, Man, and Cybernetics*, Oct 2020.
- [74] International Organization for Standardization, "ISO/IEC AWI 30173: Digital Twin — Concepts and terminology," 2021-06-08. [Online]. Available: <https://www.iso.org/standard/81442.html>
- [75] —, "ISO/IEC AWI 30172: Digital Twin — Use cases," 2021-06-08. [Online]. Available: <https://www.iso.org/standard/81578.html>
- [76] —, "ISO/FDIS 23247-1: Automation systems and integration — Digital twin framework for manufacturing — Part 1: Overview and general principles," 2021-06-08. [Online]. Available: <https://www.iso.org/standard/75066.html>
- [77] —, "ISO/FDIS 23247-2: Automation systems and integration — Digital twin framework for manufacturing — Part 2: Reference architecture," 2021-06-08. [Online]. Available: <https://www.iso.org/standard/78743.html>
- [78] —, "ISO/FDIS 23247-3: Automation systems and integration — Digital twin framework for manufacturing — Part 3: Digital representation of manufacturing elements," 2021-06-08. [Online]. Available: <https://www.iso.org/standard/78744.html>
- [79] —, "ISO/FDIS 23247-4: Automation systems and integration — Digital twin framework for manufacturing — Part 4: Information exchange," 2021-06-08. [Online]. Available: <https://www.iso.org/standard/78745.html>
- [80] Microsoft, "Cloud Computing Services — Microsoft Azure," 2021-03-03. [Online]. Available: <https://azure.microsoft.com>
- [81] J. Santa, J. Ortiz, P. J. Fernandez, M. Luis, C. Gomes, J. Oliveira, D. Gomes, R. Sanchez-Iborra, S. Sargento, and A. F. Skarmeta, "MIGRATE: Mobile Device Virtualisation Through State Transfer," *IEEE Access*, vol. 8, pp. 25 848–25 862, 2020.
- [82] J. Santa, P. J. Fernández, J. Ortiz, R. Sanchez-Iborra, and A. F. Skarmeta, "SURROGATES: Virtual OBUs to Foster 5G Vehicular Services," *Electronics*, vol. 8, no. 2, 2019. [Online]. Available: <https://www.mdpi.com/2079-9292/8/2/117>
- [83] 5G-MOBIX, "Driving forward Connected & Automated Mobility," 2021-08-18. [Online]. Available: <https://www.5g-mobix.com/>
- [84] Automotive Edge Computing Consortium, "Driving the Vision of Network and Computing Infrastructure for Connected Car Big Data," 2021-08-18. [Online]. Available: <https://aecc.org/>

Systems (IoT in ITS), and member of four other technical committees across IEEE and SAE. His research focuses on intelligent vehicle technology, including cooperative automated driving, driver behavior modeling, and vehicular cyber-physical systems.

"Digital Twin" of human-vehicle interactions, and high-speed mobility.

Rohit Gupta received the Ph.D. degree in mechanical engineering from The University of California at Santa Barbara, where he did the research for "high speed" ship design and turbulent flow modeling of gas-liquid interfaces. He is currently a Principal Cloud Architect at Toyota Motor North America, InfoTech Labs. Prior to joining Toyota, he worked in various industries: semiconductors, telecommunications, finance, and software. His research interests include cloud and edge computing architectures, connected and autonomous vehicles, AI/ML, "Digital Twin" of human-vehicle interactions, and high-speed mobility.

Kyungtae (KT) Han (M'97-SM'15) received the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin in 2006. He is currently a Principal Researcher at Toyota Motor North America, InfoTech Labs. Prior to joining Toyota, Dr. Han was a Research Scientist at Intel Labs, and a Director in Locix Inc. His research interests include cyber-physical systems, connected and automated vehicle techniques, and intelligent transportation systems.

Akila Ganlath received the M.S. and B.S. degrees in mechanical engineering from the University of California at Riverside, in 2019 and 2016, respectively. He is currently a Principal Researcher at Toyota Motor North America, InfoTech Labs. His research interests include control systems, autonomous driving, and software-driven vehicles.

Nejib Ammar received the Ph.D. degree in electrical and computer engineering from the University of California at Davis, in 2005, and the M.Sc. and B.Sc. degrees in electrical and electronics engineering from Bilkent University, Turkey, in 2000 and 1997, respectively. Since 2004, he has been an Electrical Engineer in the areas of video compression, wireless communications, cybersecurity, and intelligent connected systems. He is currently a Principal Researcher at Toyota Motor North America, InfoTech Labs. Previously he served as the Chairman of ATSC

3.0 Waveform Ad-Hoc Group.

Prashant Tiwari received the Ph.D. degree in mechanical engineering from Rensselaer Polytechnic Institute in 2004, and the MBA degree from the University of Chicago in 2016. He is currently an Executive Director at Toyota Motor North America, InfoTech Labs. Dr. Tiwari is highly active in Automotive Edge Computing Consortium (AECC) and SAE. Prior to joining Toyota, Dr. Tiwari held several leadership positions of increasing responsibilities at GE and UTC Aerospace Systems.