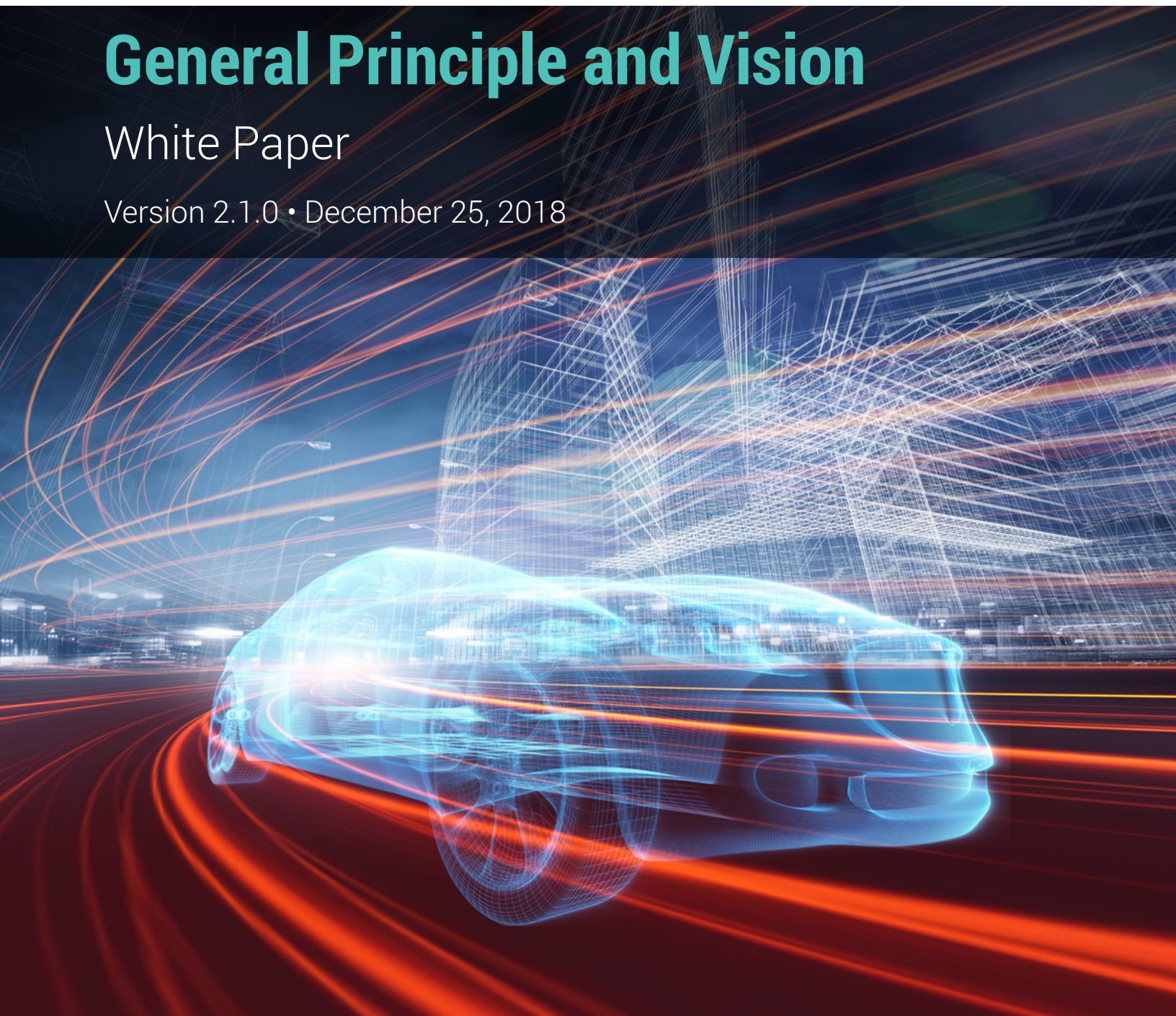# AECC

AUTOMOTIVE EDGE COMPUTING CONSORTIUM

General Principle and Vision

White Paper

Version 2.1.0 • December 25, 2018

Contents

1. Introduction

1.1. Background

To make driving safer, traffic flow smoother, energy consumption more efficient, and emissions lower, mobile communication in vehicles is increasing in importance [1][4]. Several emerging services, such as intelligent driving, the creation of maps with real-time data and driving assistance based on cloud computing, require vehicles to be connected to the cloud and networks to facilitate the transfer of a large amount of data among vehicles and between vehicles and the cloud.

The following forecasts are made for the 2025 time frame [1][2][3][4].

1. Connected vehicles will generate around US$150B in annual revenue.

2. The number of connected vehicles will grow to about 100M globally.

3. The data volume transmitted between vehicles and the cloud will be about 100 petabytes per month.

In the above forecasts, the data volume per vehicle was assumed to be only 1 gigabyte per month. This assumption was made through considering that by 2025, the only valid services are ones that can be accommodated by currently planned network capabilities and business models. Future automotive services, in fact, will transfer much larger data volumes. We estimate that the data traffic will reach from 1 up to 10 exabytes per month by 2025, at least 1000 times larger than the present volume. Based on this estimation, there will be a need for new network architectures and computing infrastructure to support massive computing resources and topology-aware storage capacity for balancing quality and cost. This, however, cannot be achieved without taking further actions, and failure to do so will limit the evolution of future services in the automotive industry.

The cellular network is one of the major mobile networks for connected vehicles, and many specifications have been standardized in the 3rd Generation Partnership Project (3GPP). However, the present work within 3GPP has not fully addressed the challenge of automotive big data, and therefore future network deployments and business models will fail to support the future needs of connected vehicles. The cellular vehicle-to-everything (C-V2X) communication considered in 3GPP, for example, mainly covers latency-sensitive safety applications and may not fully ensure the big data capacity growth between vehicles and the cloud.

Toward 5G, massive machine type communication (MTC), including narrowband (NB)-IoT has been considered by 3GPP, and it is intended to connect a massive number of small low-power sensor devices. Still, the data volumes are considered fairly modest. But adding to this, the current trend of concentrating data processing at central locations will cause huge data transmission traffic, which will lead to unnecessarily long response times and in turn will increase computation time. Assuming 20GB per month per vehicle and three million vehicles (12% market share and 25% regional ratio of 100 million vehicles), 60 petabytes of vehicle data will come to the cloud every month. Assuming the data transaction rate at the cloud is 10GB per second, it will take 70 days just for the transactions. For this reason, to be able to establish a practical platform to serve Vehicle-to-Cloud (V2Cloud) services, both computation and network performance need to be taken into account (see Figure 1).

We believe that the current mobile communication network architectures and cloud computing systems are not fully optimized to handle the requirements of connected vehicles effectively. Therefore, it is beneficial to investigate how to redesign the system architecture and reconsider network deployments to better accommodate network traffic. One possible solution is through topology-aware computing and storage resources. Our aim is to deploy this redesigned system architecture on a global scale, which will require collaboration among worldwide partners and the system architecture to comply with relevant standards.
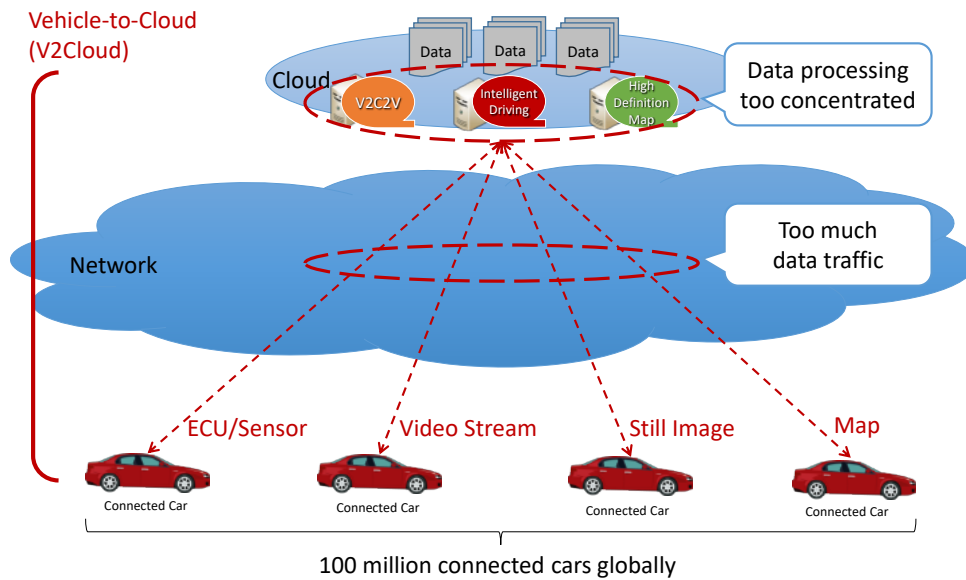
Figure 1 Problems of existing technologies' deployment

1.2. Objective

Network capacity planning has become a major challenge for mobile network operators due to the soaring costs associated with the exponential increase of data traffic. At the same time, new vertical markets, such as automotive, have an ever-increasing number of devices with high-capacity demands connected to the network, thus a new "communications offering" is needed to address these industries' specific business and technical requirements.

As next-generation networks are being standardized, we have a unique opportunity to ensure that future networks are designed and deployed to provide new services in a reasonable fashion to vertical markets such as automotive, while bringing new customers and generating new revenue for mobile network operators.

This white paper highlights the need for market actors such as communication technology companies and automotive makers to work together to ensure that future systems are designed to address the challenges mentioned above. One technology explored in the new network design will be topology-aware distributed clouds with multi-operator edge computing capabilities.

2. Concept

2.1. Distributed Computing on Localized Network

To solve the problems of data processing and traffic on the existing mobile and cloud systems described above, we introduce "Distributed Computing on Localized Networks" (see Figure 2). In this concept, several localized networks accommodate the connectivity of vehicles in their respective areas of coverage. Computation power is added to these localized networks to enable them to process local data, allowing connected vehicles to obtain responses in a timely fashion.

The concept is characterized by three key aspects:

1. **Localized Network.** A local network that covers a limited number of connected vehicles in a certain area. This splits the huge amount of data traffic into reasonable volumes per area of data traffic between vehicles and the cloud.

2. **Distributed Computing.** Computation resources are geographically distributed within the vicinity of the localized networks' terminations. This reduces the concentration of computation and shortens the processing time needed to conclude a transaction with a connected vehicle.

3. **Local Data Integration Platform.** Integration of local data by utilizing the combination of the localized network and distributed computation. By narrowing relevant information down to a specific area, data can be rapidly processed to integrate information and notify connected vehicles in real time.
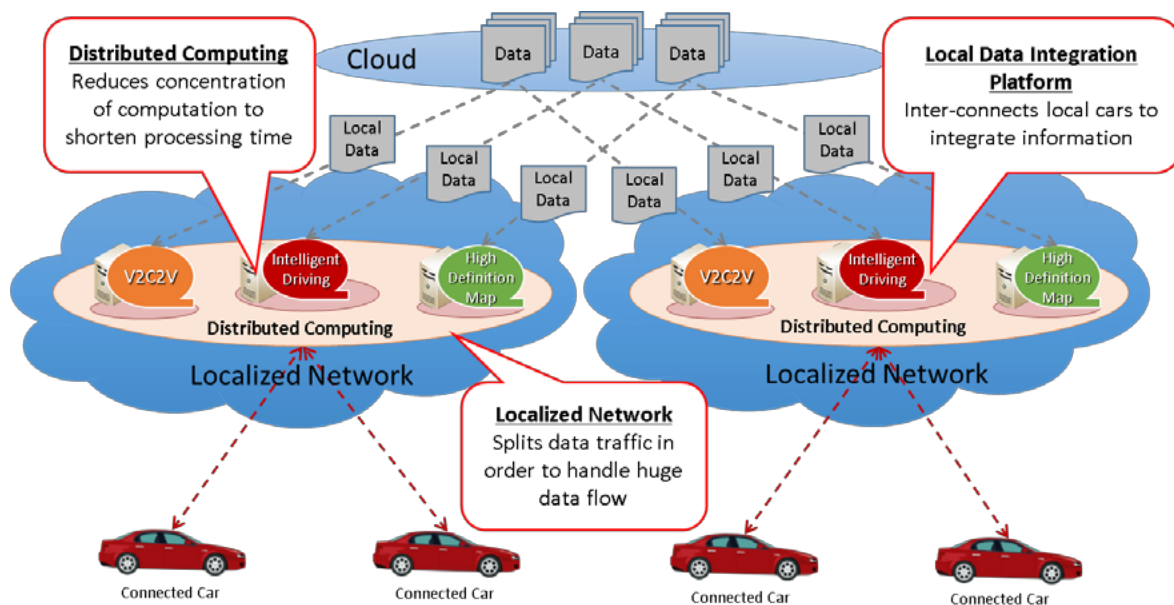

Figure 2 Distributed computing on localized networks

2.2. Edge Computing for Automotive

As mentioned in the previous chapter, the "Distributed Computing on Localized Networks" concept has three key aspects that need to be implemented, and edge computing is one promising technology, due to its features and advantages, that could be adopted to realize this concept. In the automotive use cases, edge computing technology will provide an end-to-end system architecture framework that enables distribution of computation processes over localized networks as depicted in Figure 2.

The edge computing technology used for our concept of "Distributed Computing on Localized Networks" consists of two key components: the network and the computation resources. The network is designed to split data traffic into several localities that cover reasonable numbers of connected vehicles. The computation resources are hierarchically distributed and layered in a topology-aware fashion to accommodate localized data and to allow large volumes of data to be processed in a timely manner (see Figure 3). In this infrastructure framework, localized data collected via local networks and wide area data stored in the central cloud are integrated in the edge computing architecture to provide real-time information necessary for the services of connected vehicles. In

the context of edge computing for automotive, the "edge" means the hierarchically distributed non-central clouds where computation resources are deployed, and edge computing technology can be used to design such a flexible topology-aware cloud infrastructure.
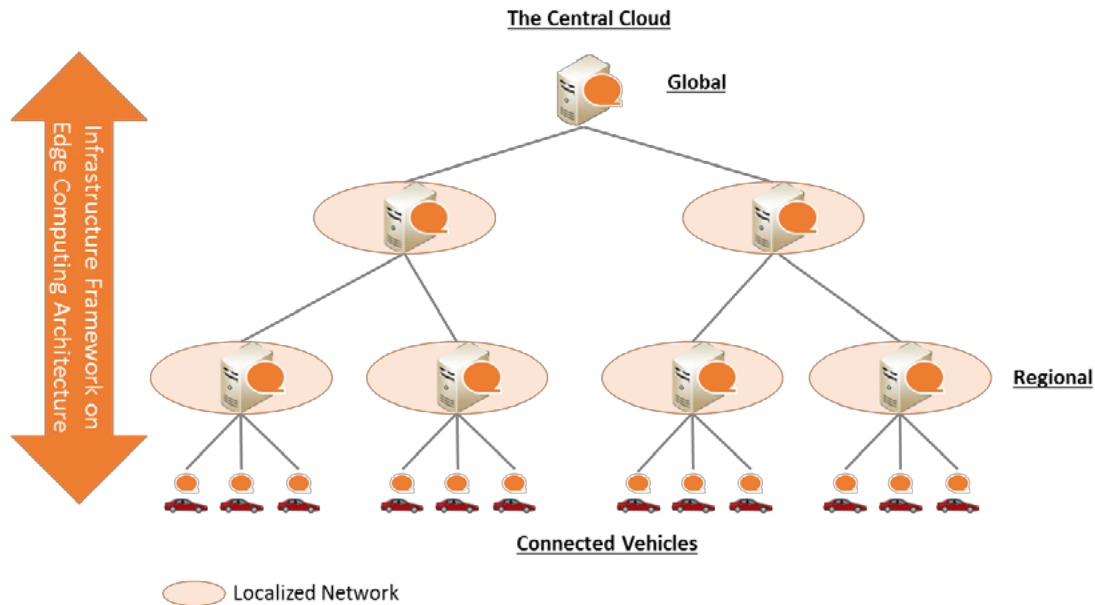

Figure 3 Edge Computing for Automotive

Edge computing is seen as the key technology to realize the "Distributed Computing on Localized Networks" concept in the automotive industry. Therefore, the Automotive Edge Computing Consortium will focus on increasing capacity to accommodate automotive big data in a reasonable fashion between vehicles and the cloud by means of edge computing technology and more efficient design of networks. The consortium will define requirements and develop use cases for emerging mobile devices, with a particular focus on the automotive industry, bringing them into standard bodies, industry consortia and solution providers. The consortium will also encourage development of best practices for the distributed and layered computing approach.

3. Service Scenarios

Network-based computation will make it possible for automotive services, especially V2Cloud services as shown in Figure 4, to come to life. These V2Cloud services cover a broad range, from sales and marketing to connected vehicle maintenance. The enhanced vehicle feature in particular is the most promising business area for next-generation connected cars. This service scenario includes, among other services, intelligent driving, high-definition map generation and V2Cloud cruising assist. These services will produce huge traffic volumes with varying levels of latency requirements.

Beyond these services, some extended services might also arise, such as telematics, insurance/financial services and traffic control. These extended services will also generate a tremendous amount of data traffic and processing for future infrastructure.
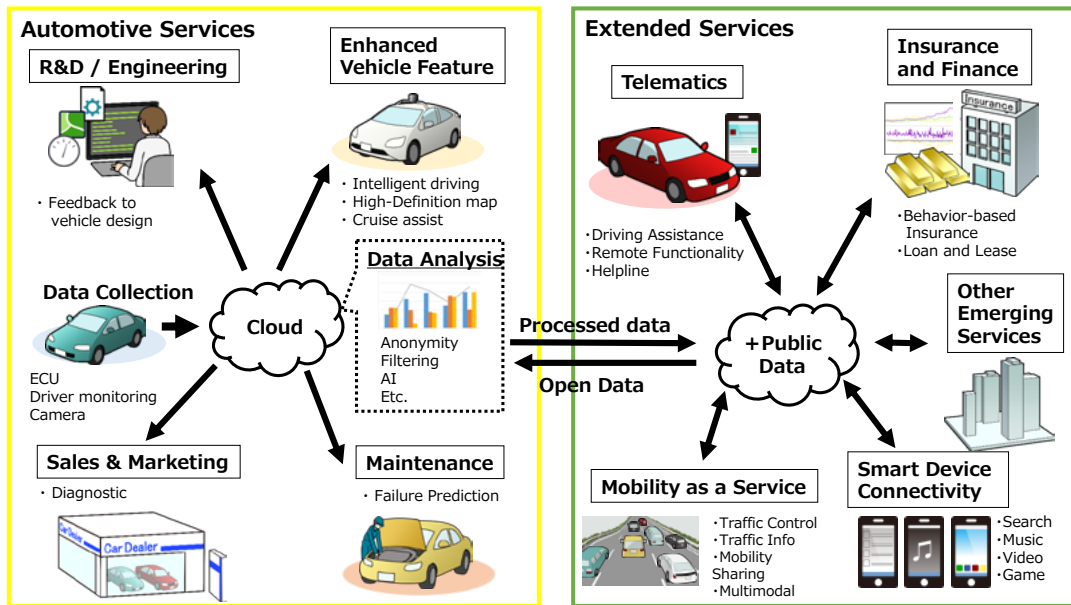
Figure 4 Emerging automotive V2Cloud services

The following examples show some typical V2Cloud service scenarios.

3.1. Intelligent Driving

Intelligent driving currently means safe and efficient driving support, but will over time incorporate autonomous driving. In this context, vehicles need to exchange vast amounts of various kinds of data with applications in the cloud. Although there are many types of service scenarios related to intelligent driving, we will describe one typical scenario for smart driving support here.

In our intelligent driving scenario, the driver's physical condition is monitored and an evaluation of his or her driving performance is given as the output. In this service scenario, the cloud service needs to collect data such as cruising data, biometric sensor data, and control data, gathered from various sources including movement logs from in-vehicle sensors and on-board biometric sensors/cameras.

The data volume is very large, creating a heavy load on both networking and computing resources and/or less optimal use of the network resources. Edge servers can help to pre-process the data on the way to the cloud, and instruct the vehicle on what data to send and how to process the data, decreasing the amount of data sent to the cloud.

The collected data is then sent to the cloud via the access networks for processing. The transfer of these data should be done effectively (preventing any loss of data) and efficiently (justifying the cost performance ratio).

To support vehicles that are moving, the AECC System shall have the ability to transfer the ongoing data session from one edge server to another.

Based on the data collected, the cloud computes the intelligent driving parameter-set using advanced machine learning techniques. It is important to send those parameters to vehicles in a timely manner. (As these parameters' data volume is relatively small, high-data volume handling is not necessary to disseminate the parameters.)

In an advanced scenario, the cloud could serve multiple vehicles via multiple cellular networks operated by different MNOs.

3.2. High-definition Map

The high-definition map consolidates static and dynamic information (e.g., vehicle position, pedestrians and obstacles) and is mandated for autonomous driving. Creating and distributing the map require many data transactions with high capacity as well as efficient processing to keep the information up to date.

This high-definition map must be able to accurately localize dynamic objects including vehicles, which is required for automated driving beyond the traditional route guidance. A large amount of data transfer is especially required to update the map. Data is collected from on-board cameras, radar sensors, and laser scanners (LIDAR), transferred and processed in the cloud. Typically, what might get sent to the cloud are deviations (Map says X, but Camera says Y). These deviations are sent to the cloud to update the high definition map.

The completed map information is stored in the center server or the edge server and needs to be distributed to relevant vehicles in a timely manner.

When vehicle is using an edge server, the edge server can provide a data compression feature to reduce amount of traffic volume between the edge server and its associated cloud. In other case, an edge server can provide a data extraction feature, for example extracting road object types and locations such as pedestrians and signals, also to reduce amount of data size transferred between the edge server and the cloud.

If network congestion happens due to concentration of vehicles, collected data to servers can be re-routed to other servers that has capacity to process it. Otherwise, the edge server experiencing network congestion can cache the collected data temporarily until the network congestion is solved.

The vehicle transportation traffic flow changes from moment to moment and sometimes create the traffic jam, which may cause the network congestion. For example, vehicle accident will increase the number of vehicles in radio cell where the accident occurs, and in turn propagates to the linked cells as well. In such a case, a lot of vehicles will upload quite similar data used for creating high-definition map. These redundant data have negative impact to the entire system. To avoid such the situation, communications traffic management scheme should be introduced.

A lot of redundant information produces huge negative impact on network and computation infrastructure as stated in above. To avoid this effects, edge server needs to identify groups of data generated from the same event such as vehicle accident by integrating them into one associated information. This kind of data integration is effective pre-processing function on edge server to reduce computation cost.

3.3. V2Cloud Cruise Assist

V2Cloud cruise assist is an example use case of a more flexible service evolution model than the conventional dedicated short-range communications (DSRC). Here the network mediates vehicle-to-vehicle communications by integrating information obtained from neighboring cars. This mechanism is called the vehicle-to-cloud-to-vehicle service or simply V2C2V. This service scenario is especially effective when used to broadcast information to vehicles that need the same information, by utilizing the combination of neighboring vehicles, roadside units and others.

Since the vehicles are expected to be in motion, the application must mediate for vehicles moving from one access point (such as a cell-tower) to another. In addition, vehicles will enter and exit the application's 'zone of interest' as a vehicle's path changes, joining or leaving a group of vehicles that are participating in the application instance. As a vehicle's journey continues, it must also be expected that a vehicle will exit one localized network and its associated edge server, entering and joining an adjacent localized network.

At first, the system needs to collect vehicle data by the same way with Intelligent Driving. But the target data includes various types of data written in the intelligent driving scenario and high-definition map scenario (e.g. cruising data, biometric sensor data, control data, on-board camera data, radar sensor data, and laser scanner (LIDAR) data).

Then, since the analyzed information can be used by vehicles in specific location (e.g. small town, cross-road, specific area of highway), an edge server can create a local information (e.g. localized high-definition map) and can update it continuously based on various collected data locally. Necessary functionality to create and manage the local information can be off-loaded to the edge server from its associated cloud.

The system to mediate data among different vehicles in the neighboring area requires fast response that meets service timing criteria (e.g. less than 10 seconds). To realize this scenario, edge server needs to be carefully selected to fulfil the criteria based on latency, availability, computation load and so on.

To realize vehicle data mediation scenario, vehicles (and in some cases local roadside units) transmit their cruising data to the cloud to be analyzed, to provide information for driving assistance (such as collision avoidance, cruise control for platooning and signal control). The system to mediate data among different vehicles in the neighboring area requires ultra-fast computing processing to fulfill service timing criteria (less than 10 seconds).

The information generated will then be distributed to relevant vehicles and roadside facilities in the neighboring area. The vehicle data distribution will require access points between the edge server, the cellular network and other (fixed line, WLAN) networks, since roadside equipment may not be attached to the edge server.

To distribute information to each vehicle, the vehicle position needs to be tracked to select a correct route to the vehicle. Especially when an edge server is used to notify a vehicle in real-time, the position in networks (incl. Cellular Network and other types of networks) is important information to identify the best route to the vehicle from the edge server.

V2Cloud Cruise Assist is a use case where a more flexible service model than most conventional services is required. The V2C2V Cruise Assist application will be executed on the edge server. Due to the nature of the application, service guarantees must be met in order to support the latency requirements. Further, during the application's lifecycle, edge server resource utilization levels will change, the application itself may require updates or maintenance, the edge server will require upgrades, etc. With a vehicle travelling at 100km/h speed vehicle covering a distance of ~ 27 meters/second, it is vital that during any operations that impact the function of the application, the impact to the vehicle is minimized.

3.4. Extended Services

3.4.1. Mobility as a Service

Many route navigation services rely on mobility data from vehicles to provide real-time navigation. The data gathered can be used by third parties to offer new services, one example being traffic flow control by road authorities. These kinds of services are the building blocks of Mobility as a Service, which will bring improvement to mobility experiences. As these services evolve, there will soon be new emerging services beyond the current ones, such as mobility sharing and multimodal navigation.

Mobility sharing is a service that includes ride sharing, car sharing and even parking lot sharing, while multimodal navigation services are end-to-end route guidance that uses various modes of transportation and also provides mobility sharing services information. Mobility sharing services will involve various types of information being shared in a timely manner between asset owners, service providers and end users; accordingly, these types of services should be built on top of intelligent driving, high-definition maps and cruise assist.

3.4.2. Finance and Insurance

Auto insurers are adopting the usage-based-insurance model by monitoring driving habits, including driving behavior, how often people drive and the times of day during which they drive. By doing so, insurers will be able to better assess the customer's risk level, which will lead to a more reasonable cost for the insurers. In a future world where real-time information can be provided to users, real-time dynamic insurance premiums will be a possible product.

Data gathered from both the vehicle, such as cruising data, and the driver's condition is processed and is delivered back to the users in the form of insurance premium information in real time. Drivers will be encouraged to drive more safely at all times, as this will lead to their eligibility for lower premiums.

The "Distributed Computing on Localized Networks" is expected to be useful for this service, as there will be a huge amount of data from several sources that must be processed quickly to be able to provide users with insurance premium information in real time.

4. Service Requirements

Given the service scenarios described in the previous chapter, service requirements will include the following parameters.

- **Data Generation and Traffic Rates.** Amount of data generated inside vehicles and amount of data transmitted between vehicles and the cloud. Vehicles are moving data sources that generate massive volumes of data, which results in heavy uplink traffic. This moving data source is characterized by its high mobility and not-always-on connectivity, which is quite the opposite of the present service requirements for smartphone and internet usage. This is the main requirement for determining the appropriate system architecture for handling the required data processing for services described in this document.

- **Response Time.** Response time between a vehicle and the cloud including deviation with regard to latency. These requirements are critical for some of the service scenarios, including vehicle control based on real-time information (such as positions of other vehicles and pedestrians).

- **Availability-Cost Tradeoff.** Some services need less network availability; as a result, cost -effectiveness can be prioritized. Other services, on the other hand, require full cloud service availability regardless of the cost. These considerations call for more diverse network options to balance availability and cost.

- **Data Security and Privacy.** Some of the expected service scenarios include highly confidential data that must be secured to maintain privacy and security. This mandates that the distributed network honor such requirements, with solutions that can give the appropriate security level while keeping service reliability.

- **Data Locality and Data Sovereignty.** The service needs to align with the rules and regulations regarding data locality and data sovereignty where the data is collected and processed. Compliance requirements for data hosting differ among countries. Depending on these requirements, data locality might differ between services and locations.

- **Service for Multiple Vehicle Systems via Multiple Cellular Networks.** In a future scenario, the cloud could be operated by the Mobility Service Provider (MSP) and could serve multiple vehicle systems via multiple cellular networks operated by different MNOs.

Table 1 shows the necessary requirements per service scenario.

System Requirements *		V2Cloud Cruise Assist	High-definition Map Generation & Distribution	Intelligent Driving
Major Data Source		Video Stream	Still Image (road surface image)	ECU data
Data Generation in vehicle		~ 1215EB/month[1]	~ 375EB/month[2]	~ 22.5EB/month[3]
Target Data Traffic Rate		1 ~ 10EB/month in total (cost constraint might limit this number)		
Response Time	Uplink	< 10 seconds	< 1 week	< 1 week
	Downlink	< 10 seconds	< 1 week	< 10 minutes
Required Availability	Uplink	Continuous	Occasional	Occasional
	Downlink	Continuous	Occasional	Continuous

Table 1 System Requirements

* The numbers in Table 1 are total values for 100 million connected cars.
[1] [Preliminary assumption] Video stream: 10Mpixel*3Byte(Color)*1/4(Lossless JPEG)*30fps, Average travel time: 30 min/day
[2] [Preliminary assumption] Still image: 10Mpixel*3Byte(Color)*1/4(Lossless JPEG) at every 2 meters, Average travel distance: 1000km/month
[3] [Preliminary assumption] Automotive Ethernet: 100Mbps*1/3(effective), Average travel time: 30 min/day

As indicated in the above table, some of the predicted performance requirements will be difficult for the current communication infrastructure to manage. Note that it will count more data from laser scanners for outside situational awareness, known as LIDAR. Therefore, it is important to discover any missing links in the technology and to find out how the technology is being deployed in order to realize the envisioned service scenarios, by analyzing the gap between the desired requirements and the existing technology and deployments.

5. Next Steps

This consortium will investigate cutting-edge technologies to fulfill the system requirements described in the previous chapter. These technologies should include flexible topology-aware distributed clouds with multi-operator edge computing capabilities, appropriate AI-enabling technologies, improved radio access technologies and other needed technologies. We aim to reveal the best practices in combining these potential technologies to create a provisional reference architecture for next-generation connected vehicles (see Figure 5).
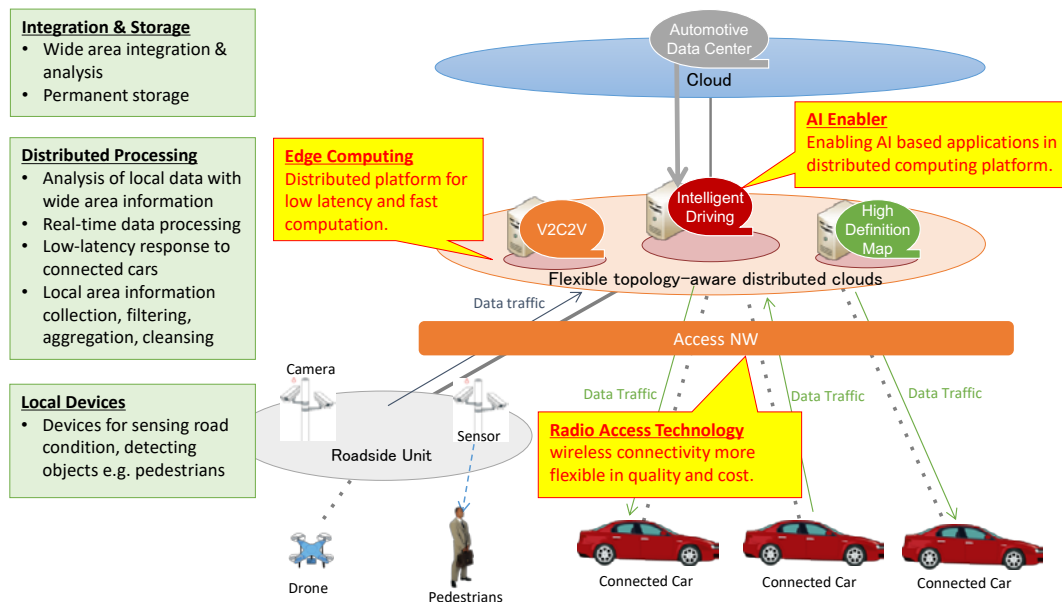
Figure 5 Potential Technologies

- **Edge computing.** Here, the computation resources are moved away from the central datacenter to be distributed further out in the networks, which means the hierarchically distributed non-central clouds where computation resources are deployed. As mentioned, edge computing technology is defined as distributed and even layered computing technology with localized networks, which involves challenges in both computing and networking. Our focus will be to ensure that the network infrastructure can be utilized to improve the characteristics of the indicated services, including the realization of real-time application response through a low-latency network environment and distributed computing.

- **AI enabler.** Artificial intelligence technologies such as machine learning will implement required intelligence capabilities to support autonomous driving, cruising assist, creation of high-definition map information, etc., which requires big data and highly intelligent analysis. Our focus will be on technologies enabling such AI-driven services in distributed computing with localized networks.

- **Radio Access Technologies.** Wireless technologies will be used to connect a vehicle with distributed computing platforms with more flexibility in quality and cost. This includes not only cellular technologies but also local radio access such as Wi-Fi and low-power wide-area (LPWA).

- **ID Management for Multiple Vehicle Systems via Multiple Cellular Networks** (To be described in the next version)

- **Mobility Support**. (To be described in the next version)

- **Data Transfer Preference.** (To be described in the next version)

This investigation will help us in deciding the necessary architecture and deployment to realize a distributed cloud for automotive based on the expected requirements for each service scenario and the technology concepts stated in this document.

The consortium will produce a strategic roadmap to introduce these new technologies to the existing infrastructure in order to realize our future vision. The roadmap will cover various aspects, including technology deployment as well as appropriate business schemes and charging models, and multi-operator situations.

6. Summary and Conclusions

Network-based computation will make it possible for the next generation of automotive services to come into being. The expected service scenarios include intelligent driving, high-definition map generation, V2Cloud cruising assist, etc. Autonomous vehicle services, which will require huge traffic volumes and low latency, will require flexible topology-aware distributed clouds with multi-operator edge computing capabilities.

In the concept of distributed computing on localized networks, several localized networks accommodate the connectivity of vehicles in their respective areas of coverage. Computation power is added to these localized networks so that they are able to process local data, enabling connected vehicles to obtain responses in a timely fashion. To realize the flexible topology-aware distributed clouds, edge computing is a key technology. For automotive use cases, edge computing technology will provide an end-to-end system architecture framework used to distribute computation processes from centralized networks to localized networks.

The Automotive Edge Computing Consortium will focus on increasing capacity to accommodate automotive big data in a reasonable fashion between vehicles and the cloud by means of edge computing technology and more efficient design of networks. The consortium will define requirements and develop use cases for emerging mobile devices, with a particular focus on the automotive industry, bringing them into standard bodies, industry consortia and solution providers. The consortium will also encourage development of best practices for the distributed and layered computing approach.

7. Terms and Definitions

Term	Definition
Cloud	The concept of cloud computing, which is network-based computing providing processing resources and data on demand
Central Cloud	Top-most processing and storage resource in the architecture
Connected Vehicle	A vehicle equipped with network access to share vehicle data with other devices and servers via a network
Cruising Data	The data collected from a client vehicle regarding its movement
Data Locality	Where and how data should be stored and processed in the cloud space
Data Sovereignty	The handling procedures for data in accordance with the local jurisdiction's requirements
Distributed Computing	A model in which computing nodes are networked so as to communicate and coordinate in order to achieve a certain goal
Edge Computing Technology	Distributed and evenly layered computing technologies within the localized network, which is composed of both computing and network aspects
Flexible Topology Aware Distributed Cloud	A cloud solution that executes applications in a topology and geographically aware fashion, which means that the topology can be determined based on application requirements and the capability of the cloud instances to execute the application and handle its related data, according to the required cost and quality balance
High-definition Map	Detailed map for vehicle driving including dynamic information of the road, objects, pedestrians and so on
Intelligent Driving	Intelligent driving support (including autonomous driving) with AI technologies
Local Data Integration Platform	The platform that integrates data on the localized network and the distributed computation
Localized Network	A local network that covers a limited number of connected vehicles in a certain area
Multi-operator	Operators of network and computing platforms, which includes telecom operators and enterprises
Telematics	Services that involve telecommunications and information processing
V2Cloud	Communication between the vehicle and the cloud

8. References

[1] "A Business Case for Connecting Vehicles", Cisco (2011).
https://www.cisco.com/c/dam/en_us/about/ac79/docs/mfg/Connected-Vehicles_Exec_Summary.pdf

[2] "2025 Every Car Connected", GSMA (2012).
https://www.gsma.com/iot/wp-content/uploads/2012/03/gsma2025everycarconnected.pdf

[3] "Connected Car Global Forecast 2015", SBD (2015).
https://www.sbdautomotive.com/files/sbd/pdfs/536%20connected%20car%20forecast%20ib%2015.pdf

[4] "Connected car report 2016", PwC (2016).
https://www.pwc.nl/en/publicaties/connected-car-report-2016.html