Automotive Edge Computing Consortium

General Principle and Vision

White Paper

Version 1.0.0

December 1st, 2017

Table of Contents

This is the first version of the white paper, which will serve as input to the consortium activities. The white paper will be continuously updated during these activities.

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 around 100M globally.
3. The data volume transmitted between vehicles and the cloud will be around 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 by considering that around 2025, the only valid services are the ones that can be accommodated by the currently designed network capabilities and business models. The future automotive services, in fact, will require much larger data transfer capacity. We estimate that the data traffic will reach 10 exabytes per month around 2025, approximately 10,000 times larger than the present volume. Referring to this estimation, there will be a need for new network architectures and computing infrastructure to support massive computing resources and topology-aware storage capacity in terms of 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 present work regarding automotive Internet of Things (IoT) in the 3rd Generation Partnership Project (3GPP) has not fully addressed this challenge, and the present network deployments and business models do not yet support the future needs of the connected vehicles. The cellular vehicle-to-everything (C-V2X) communication considered in 3GPP mainly covers latency-sensitive safety applications and may not fully ensure the big data capacity growth between vehicles and the cloud. Massive machine type communications (MTC), including narrowband (NB)-IoT, in 3GPP is intended to connect a massive number of small low-power sensor devices. Still, the data volumes are considered to be fairly modest. Adding to this, the current trend of concentrating the data processing at central locations will cause huge data traffic transmission, which leads to unnecessarily long response times and in turn will increase the computation time. For this reason, to be able to establish a practical platform to serve Vehicle-to-Cloud (V2Cloud) services, both the computation capacity and the network performance need to be taken into account (see Figure 1).

We believe that the current mobile communication network architectures and conventional cloud computing systems are still not fully optimized to handle the requirements of the connected vehicles effectively. Therefore, it is beneficial to investigate how to redesign the system architecture and reconsider the network deployments to better accommodate the 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 the relevant standards.

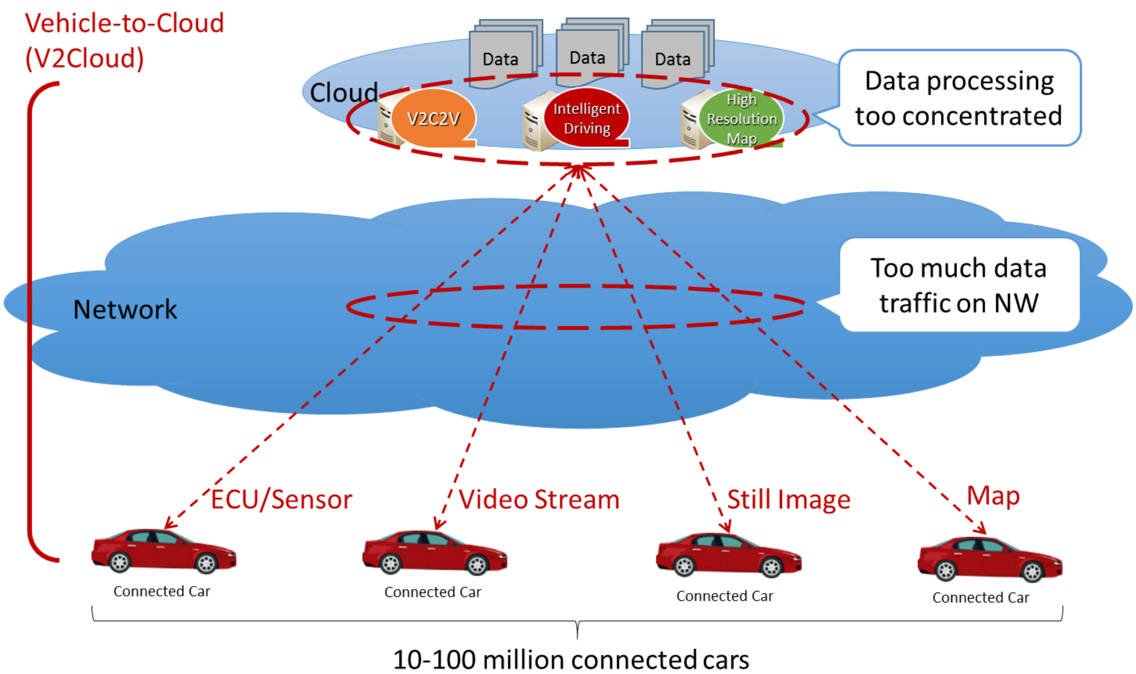

Figure 1 Problems of Existing Technologies' Deployment

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 networks 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 Network” (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 be able to process local data, enabling the 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 where the localized networks are terminated. 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 the 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.

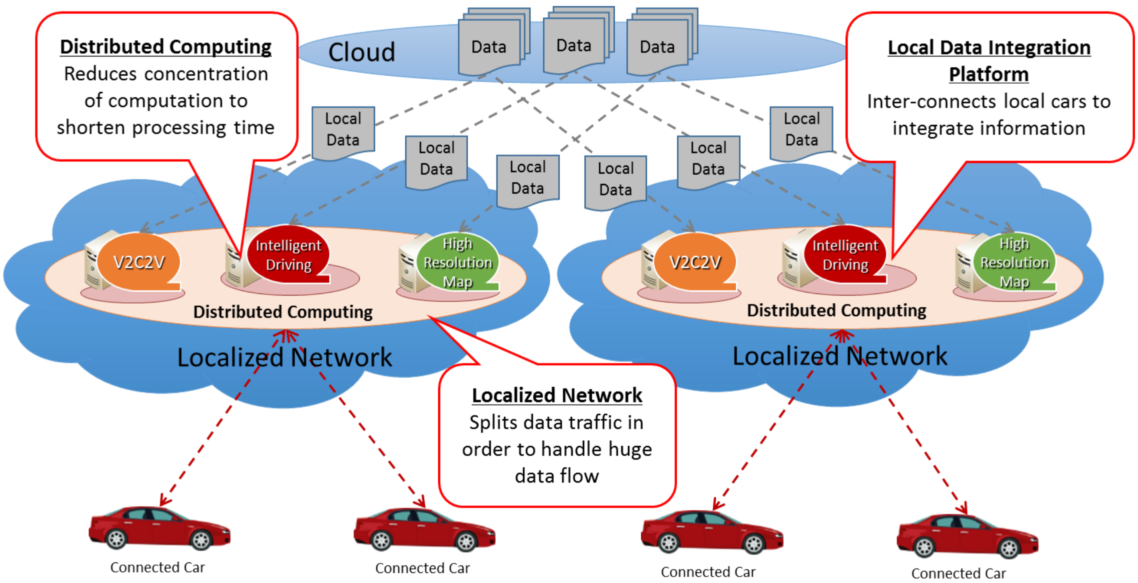

Figure 2 Distributed Computing on Localized Network

2.2. Edge Computing for Automotive

As mentioned in the previous chapter, the “Distributed Computing on Localized Network” concept has 3 key aspects that needs to be implemented, and edge computing technology is one promising technology in terms of its features and advantages to be adopted to realize this concept. In the automotive use cases, the edge computing technology will provide an end-to-end system architecture framework that enables to distribute computation processes over localized networks as depicted in Figure 2.

The edge computing technology used for our concept of “Distributed Computing on Localized Network” 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 volume 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 on the edge computing architecture to populate real-time information necessary for 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 the edge computing technology can be used to design such a flexible topology-aware cloud infrastructure.

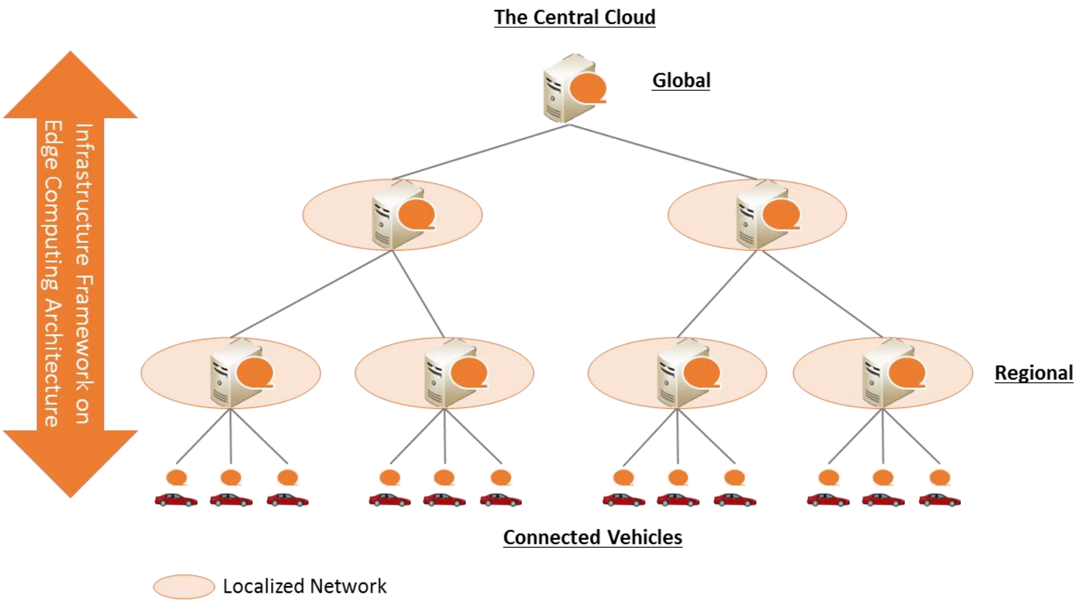

Figure 3 Edge Computing for Automotive

The edge computing is considered as the key technology to realize the “Distributed Computing on Localized Network” 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 consortiums, 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 the automotive services, especially V2Cloud services in Figure 4, to come to life. These V2Cloud services cover a broad range, from sales and marketing to connected vehicles 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 resolution map generation, and V2Cloud cruising assist. These services will produce huge traffic volumes with varying grade 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 generate a tremendous amount of data traffic and processing for future infrastructure.

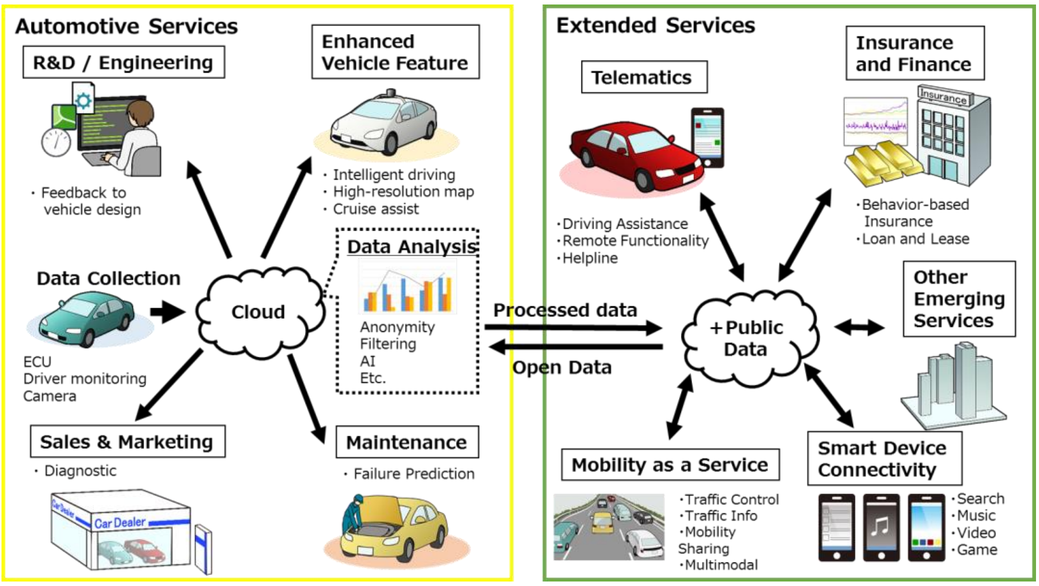

Figure 4 Emerging automotive V2Cloud services

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 amount of various 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 the driving performance is given as the output. In this service scenario, data such as cruising data, sensor data, and control data need to be collected and heavily processed, which involves machine learning. The data is gathered from various sources including movement logs from in-vehicle sensors and on-board biometric sensors/cameras and is transferred to the cloud to be processed. Real-time response from the cloud is important. “Distributed Computing on Localized Network” concept is a possible infrastructure enabler for this service scenario.

3.2. High Resolution Map Creation and Distribution

The high-resolution map consolidates static and dynamic information (i.e. vehicle position, pedestrians, and obstacles) and is mandated for autonomous driving. Creating and distributing the map requires a lot of data transactions with high capacity and in certain cases low latency, as well as reliable communication and efficient processing to keep the information up to date.

This high-resolution 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 and real-time processing is especially required for the map’s dynamic data. Data is collected from on-board cameras and laser scanners (LIDAR) and transferred and processed in the cloud. The completed map information also needs to be distributed to relevant cars in a timely manner.

The "Distributed Computing on Localized Network" concept is a possible infrastructure model for this service scenario.

3.3. V2Cloud Cruise Assist

V2Cloud cruise assist is one example use case of a more flexible service evolution model aside from the conventional dedicated short-range communications (DSRC). Here the network mediates the 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.

To realize this service 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, cruising control for platooning and signal control). The generated information will then be distributed to relevant vehicles and roadside facilities in the neighboring area. This kind of transport system among different vehicles in the neighboring area through the network requires low-latency communication and ultra-fast computing processing to fulfill the service timing criteria. The "Distributed Computing on Localized Network" concept is considered to play an important role to support these use cases.

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 gathered data can be used by third parties to offer new services, one example being the 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 experience. 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 which includes ride sharing, car sharing, and even parking lot sharing, while multimodal navigation services is an end-to-end route guidance that uses various modes of transportation and also provides mobility sharing services information. Mobility sharing services will involve various information being shared in a timely manner between asset owners, service providers and end users, thus, these types of services should be built on top of intelligent driving, high resolution map, 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 they 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 world where real time information can be provided to users, real-time dynamic insurance premium will be a possible product of the future.

Data gathered both from the vehicle, such as cruising data and driver's condition is processed and is reflected to the users in the form of insurance premium in real time. The driver, will be encouraged to drive safer at all time, as this will lead to their eligibility for a lower premium.

The "Distributed Computing on Localized Network" is expected to be valid for this service, as there will be a huge amount of data from several sources which must be processed quickly to be able to provide the users with their insurance premium 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 volume of data, which results in a heavy uplink traffic. This moving data source is characterized by its high mobility and not-always-on connectivity, which is quite an opposite from the present service requirements for smartphone and internet usage. This is the main requirement to determine the appropriate system architecture serving the required data processing of services described in this document.

- **Response Time**

  Response time between a vehicle and the cloud including the deviation in terms of 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, thus cost efficiency can be prioritized. Other services, on the other hand, require full cloud service availability regardless of the cost. This situation calls 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 maintained privacy and security. This requires that the distributed network honor such requirements, with solutions that can give the appropriate security level while keeping the reliability of the service.

- **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.

Table 1 shows the necessary requirements per service scenario.

Table 1 System Requirements

System Requirements *		V2Cloud cruise assist	High-resolution 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		**~ 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**

* - The numbers in Table 1 are total values for 100 million connected cars.

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 scanner for outside situational awareness, known as

[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

LIDAR. Therefore, it is important to discover any missing links in the technology and to find out how the technology is being deployed to realize the envisaged 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 practice in combining these potential technologies to create a provisional reference architecture for next-generation connected vehicles (see Figure 5).

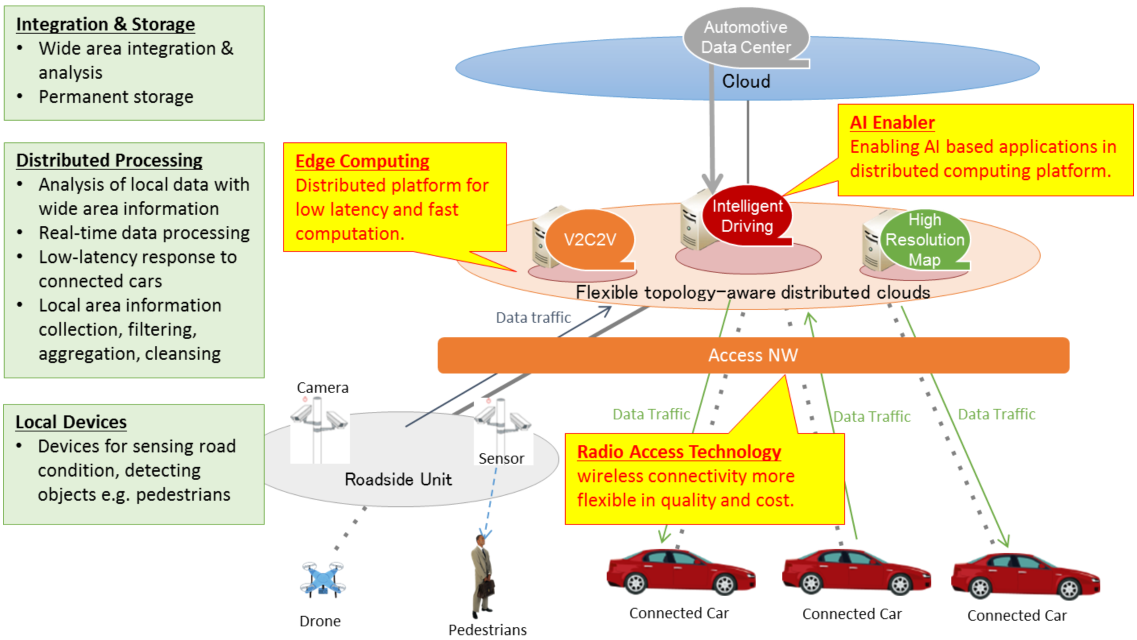

Figure 5 Potential Technologies

- **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 network, which involves both challenges in 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 to support autonomous driving, cruising assist, creation of high-resolution 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 network.
- **Radio Access Technologies**. Wireless technologies will be used to connect a vehicle with distributed computing platform 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).

This study 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 to realize our future vision. The roadmap will cover various aspects including technologies deployment as well as appropriate business schemes and charging models, including multi-operator situations.

6. Summary and Conclusions

Network-based computation will make it possible for the next generation of automotive services to come to life. The expected service scenarios includes intelligent driving, high resolution map generation, and V2Cloud cruising assist, etc. The 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 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 be able to process local data, enabling the connected vehicles to obtain responses in a timely fashion. To realize the flexible topology-aware distributed clouds, edge computing is considered to be 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 consortiums, and solution providers. The consortium will also encourage development of best practices for the distributed and layered computing approach.

7. Appendix

7.1. Terms and Definitions

Term	Definition
Cloud	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 procedure of a data in accordance to the local jurisdiction.

Distributed Computing	A model in which computing nodes are networked to be able 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 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 the topology can be determined based on the applications 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-resolution Map	Detailed map for vehicle driving including dynamic information of road, objects, pedestrians, and so on.
Intelligent Driving	Intelligent driving support (incl. autonomous driving) with AI technologies.
Local Data Integration Platform	The platform which 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 which involves telecommunications and information processing.
V2Cloud	Communication between the vehicle and the cloud.

8. Contributors

DENSO Corporation (www.denso.com)

Tomohiro Miyachi

Ueno Takemi

Hidenori Akita

Ericsson (www.ericsson.com)

Christer Boberg

Per Synnergren

Toshikane Oda

Intel Corporation (intel.com)

Soo Jin Tan

Daouda Fall Ndiogou

Nippon Telegraph and Telephone Corporation (https://www.ntt-global.com/)

Koya Mori

Lidwina Andarini

NTT DOCOMO, INC. (https://www.ntt-global.com/)

Atsushi Minokuchi

Toyota InfoTechnology Center Co., Ltd. (https://www.toyota-itc.com/en/)

Ryokichi Onishi

Toyota Motor Corporation (http://newsroom.toyota.co.jp/en/)

Kenichi Murata

Kiichi Iwasaki

9. References

[1] “A Business Case for Connecting Vehicles”, Cisco (2011).

[2] “2025 Every Car Connected”, GSMA (2012).

[3] “Connected Car Global Forecast 2015”, SBD (2015).

[4] “Connected car report 2016”, PwC (2016).