MAD - Urban -

AP 1.2 Architecture

Version 1.4

(Working Draft for EUCAD 2024)

Authors

Julian Schindler, DLR

Jürgen Weimer, DLR

Tobias Fleck, FZI

Philip Schörner, FZI

Manuel Osebek, DLR

Sebastian Scheibe, DLR

Bernd Gassmann, Intel

Florian Mirus, Intel

Reza Dariani, DLR

Tobias Birkle, GFT

Table of content

1 Goal of the Document

Managed Automated Driving (MAD) is a new approach for automated driving and an extension or alternative for established Connected Cooperating Automated Driving (CCAM). The goal of the document is to draft the system architecture for MAD as basis for discussion with international actors in the field of automated driving. (Operators, Users, Vehicle makers, Government, Standardization Bodies, Public Private Partnerships like CCAM, 2 Zero, ERTRAC, ERTICO etc.).

The document will first provide in section 2 the motivation for MAD and an overview about the MAD concept with its relation to national and international developments outside of the project. Then, section 3 gives an overview about the overall system architecture, also introducing important main components. These components are described in detail in section 4, including their interfaces. Afterwards, section 5 describes all functions which make use of several components and their interrelation. Therefore, also the data exchange using the interfaces is described here. Finally, section 7 describes the example implementations within the project MAD Urban of the instances at the different test sites.

Version 1.0: Initial draft of sections 1-4. It will be iterated through the project.

Version 1.4: Working Draft, Branch for EUCAD 2025

Version 2.0: Target: Whitepaper

2 Overview MAD

2.1 Motivation

The concept of Managed Automated Driving has emerged from the specific requirements of an automated modular vehicle concept (U-Shift) and a strict, business model-driven approach.

In this context, it has been recognized that virtually all known automated vehicles are confronted with problems that cannot be solved at all with current vehicle-based automation solutions, or only to a very limited extent. Three key aspects are e.g.:

- **Safety:** Automated vehicles do not master complex traffic situations in urban space with the expected safety. The vehicles detect only part of the objects, especially occluded objects which might be vulnerable road users (VRU). Avoidance strategies with very low driving speed in cities are not economic for operators and not user-friendly. Ignoring the risk of a fatal incident with a VRU could have dramatic impact to the business of operators.
- **Profitability:** Automation needs to have several sensors for all possible situations and conditions while the sensor and automation components like computation units (CU) for perception are expensive. With a large scale roll out of automated vehicles, investors and operators have to think where to invest. Cost for vehicle technology increase strongly with the number of vehicles in operation. Cost for infrastructure have a high entrance but are nearly independent of number of vehicles in operation. At a certain number of vehicles, the invest in infrastructure pays well.
- **Energy:** Sensors and required components for executing automation functions in the vehicles consumes a high amount of energy and induce more weight to the vehicles. The efficiency chain of electric driven automated vehicles worsens this high energy demand. A proportional increase in automated vehicles with automation technology in the vehicle would therefore be harmful to the climate. MAD with automation in infrastructure

doesn't have the problem with the efficiency chain of battery systems, has a better usability of the perceptions systems on large scale and energy losses could be partially used for heating.

Overall a transformation from vehicle-based automation to an infrastructure-based automation is recommended. The transformation should be a continued way from today's CCAM to MAD, stepwise shifting automation functions into the infrastructure. The application focus are cities and urban regions with mostly complex traffic situations. It is open for all AVs, also such used for interurban travel and transport entering MAD controlled areas in cities.

2.2 The MAD Basic Principle

Full MAD is characterized by two main features:

1) MAVs (Managed Automated Vehicles) are completely controlled via an edge control unit (ECU) and the backend – In the full MAD version, the majority of the automation functionality migrates to the infrastructure (world-model, decision&control semantic understanding, etc.) to the point that the vehicle becomes a simple actuator (with necessary safety functionality).
2) Complete sensory coverage of the urban operational area (mostly: roads) via a set of infrastructure sensors. In the ideal MAD configuration, corresponding Road Capturing Units (RCUs) are set up along the entire route in order to enable comprehensively safe and optimized driving operation even in complex road situations.

Besides full MAD, a full spectrum of infrastructure support is possible when RCUs are available. Vehicles can just be informed by infrastructure about e.g. detected objects (as done in collective perception), advice about optimal behavior can be provided, etc. This offers a valid migration path from less cooperative approached up to full MAD, esp. by using existing protocols e.g. by ETSI also in MAD.

2.3 Today's Roadmaps

Today different Roadmaps show how future automation should be used for accident free road transport or better traffic flow.

The **CAR 2 CAR** Communication Consortium (C2C-CC) aims at assisting towards accident free traffic (vision zero)[1]. They are focused on communication and their idea is based on sharing information, awareness, perception and intentions. Key automation stack is in vehicles and their cooperation addresses other vehicles and infrastructure (Fig.1). The consortium provides a very good basis especially in the field of communication and the necessary standardisation as well as its progress in Public Key Infrastructures (PKIs). CAR 2 CAR is still active but although their good work there is no visible breakthrough in automated driving. The key weakness of CAR 2 CAR idea is that the cooperative V2X data exchange works only when users come closer to each other. The safety problem cannot be solved simple because density of other partners and or infrastructure is to low and roll-out or mi ration is not sufficiently addressed.

The **ERTRAC-CCAM-Roadmap**[2] discusses the Levels of Infrastructure Support for Automated Driving (ISAD) as vision for potentially exceeding the ODD and so also contributing to road safety and efficiency. It is in line with the CAR 2 CAR approach and it's day-x scenarios (Fig.2). ISAD has been refined during the latest versions and discussions about business models have been

[1] https://www.car-2-car.org
[2] ERTRAC-CCAM-Roadmap-V11.pdf

included. However, the current focus is on autobahnen with incident detection, road-works or road-sign traffic management.

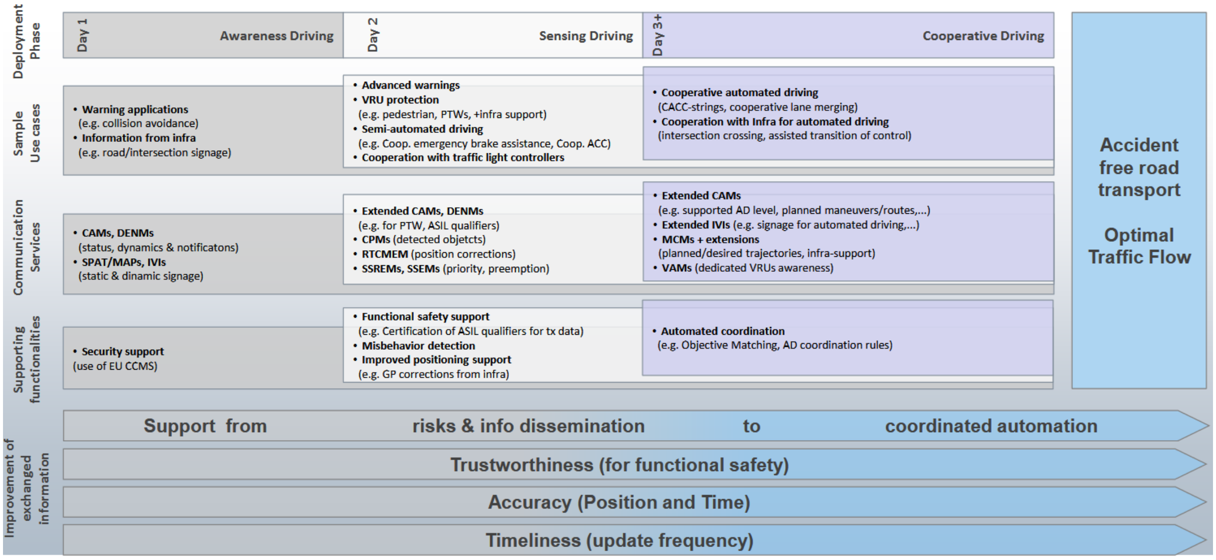


Figure 1: The Car-2-Car Roadmap showing the way to cooperative driving[3]

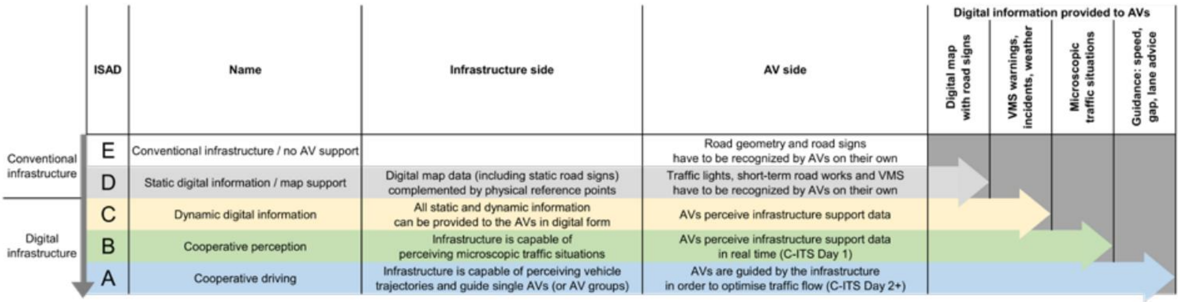


Figure 2: Levels of Infrastructure Support for Automated Driving (ISAD) [4]

Both consortia or PPPs are strong established but not consequent. Both see the advantages of infrastructure but do not consequently claim the fundamental need of a large-scale introduction of digital infrastructure, especially in complex urban traffic situations. Furthermore, C2C and ISAD scenarios end all only on a cooperation level "Day 3" with recommendations or soft guidance commands. This is also not consequent. For traffic and operational optimization, a strong management with control commands are necessary.

MAD will be based on todays C2C or Ertrac work. But MAD vision goes beyond these two automation-strategy examples and claims the massive expansion of digital infrastructure for automated driving and a consequent management platform for traffic optimization and safety. Infrastructure investments will pay of better if number of users is not limited to single domains.so MAD has to comprise all application domains and their specific needs and interfaces from Public Transport, Logistics or MIV to operations in confined areas.

[3] *https://www.car-2-car.org/fileadmin/downloads/PDFs/roadmap/Roadmap_2020_figure.pdf*

[4] *ERTRAC-CCAM-Roadmap-V11.pdf page 25*

2.4 Cooperation and Connectivity

Currently, most vehicles do not have any cooperative features and only low-level vehicle automation functionality. Nevertheless, both aspects are addressed by the industry, resulting in rising available automation levels (today: SAE level 3) and first cooperative driving functions on the market. In terms of communication, two trends can be seen:

On the one hand local communication such as ITS-G5 is used which allows several local cooperation measures, as depicted in Figure 1. The measures start from simple state information incorporated in Day-1 messages. In Day-2 messages, mostly sensing is in focus, e.g. collective perception. Finally, Day-3(+) messages also include real cooperative driving functions, such as maneuver coordination etc.

On the other hand, cloud communication is also getting in focus. Besides the entertainment functionality in vehicles, such systems are also used for data acquisition by the respective OEMs to gain insights into current driving (and driver) situations, to use the data for e.g. AI training approaches for automated driving and to offer extended services to the passengers of vehicles.

Most likely, both types of communication will be used in the future in a combined way.

The MAD approach also takes these developments into account. Since road infrastructure and esp. road capturing units (**RCUs)** are running on the road, the provided information can also be used in line with ITS-G5 Day-2 and Day-3(+) messages. The general idea is that all infrastructure-sensed objects are included in CPMs to enhance the perception in receiving entities. In addition, also behavioral information such as lane and speed advisories and preferred paths and maneuvers will be provided using e.g. extended CAMs, extended IVIs or MCMs. By adhering to the existing standards and by also taking part in future standardization activities at C2C-CC and ETSI a given backward-compatibility is guaranteed, which also supports the MAD deployment on the road.

In detail, the following vehicle types regarding communication and automation capabilities are addressed:

- Managed Automated Vehicles (MAV): These vehicles are fully controlled by the infrastructure on specific levels. MAVs are designed to work only in areas where the infrastructure is providing the required level of support. MAVs therefore will have a reduced set of on-board sensors and computing power. Most likely, first MAVs therefore are fleet vehicles performing specific tasks (e.g. hub-to-hub logistics, fixed-route public transport).
- Cooperative Automated Vehicles (CAV): For backward compatibility, all connected/cooperative automated vehicles are also benefiting from broadcasted standard messages. Since only the communication part is affected, the possible level and kind of the automation function is up to the respective OEM or automation function provider.
- Connected Vehicles (CV): Also, for backward compatibility, non-automated vehicles equipped with the corresponding communication hardware and the respective HMI are also benefiting from the broadcasted standard messages. Of course, the impact of the messages will be limited since a human driver needs to transfer the message content into the appropriate actions.
-

3 MAD Overall System Architecture

Managed Automated Driving requires several technical components which need to be aligned. This chapter provides an overview on the full system architecture and some general terms required for understanding the general concept and the interrelation of the components.

First, the general MAD system topology is shown to provide the overview. Since the road side infrastructure of MAD is deployed locally on the road, it is also important to take a look at the geometrical aspects of the MAD system. These aspects are described in the second sub section showing the MAD spatial topology.

3.1 MAD System Topology

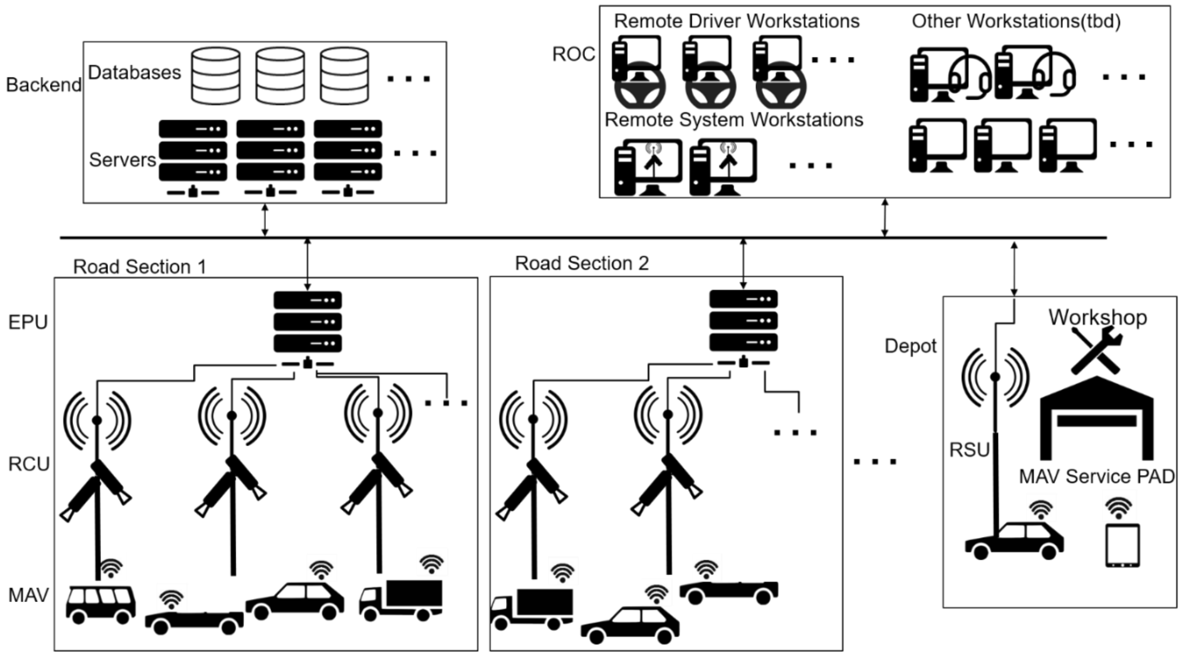


Figure 3: MAD System Overview

As shown in Figure 3, an MAD system is including several components. These are:

- Vehicles: Different kinds of vehicles with different cooperation capabilities can take part in the MAD approach. The Managed Automated Vehicle (MAV) is representing the full control by the infrastructure. More details about the vehicle types can be found section
- Road Capturing Unit (RCU): This is a unit of sensors (e.g. a pole) able to monitor a drivable section of the world, e.g. a part of a road or intersection. It is described in more detail in section *4.2 Road Capturing Unit*.
- Edge Processing Unit (EPU): The information of the different RCUs is integrated locally in road side edges. An EPU therefore takes over the fusion of this information and the provision of locally relevant advisories or – in case of MAVs – also vehicle control commands. More details can be found in section *4.3 Edge Processing Unit*.
- Backend: Different EPUs are controlled and monitored in a backend. This entity also is responsible for logging of data, esp. detected objects and related actions of the EPUs.
- Customer API: The data of the whole system can be made available also to external customers, e.g. to related city, to road operators etc. Also, the inclusion of external data by the system, e.g. temporary traffic management measures, traffic light states, weather data, destinations for on-demand MAVs etc. needs to be fed into the system. This is

done using Costumer APIs. More details about those can be found in section *4.5 Customer API*.

- MAD Operation Center (MOC): The supervision of the whole system is done in the MOC. For MAVs this also includes the supervision of the MAVs currently available in the monitored area. As these vehicles are basically SAE Level 4 vehicles, also the possible remote assistance or even remote control is done in dedicated sections of the MOC. More details can be found in section *4.6 MAD Operation Center*.
- Depot: For servicing single MAVs or MAV fleets, one or more depots and/or workshops are required. A depot requires specific handling, like a direct remote control of a MAV by depot personnel. More details about the depot and the related sub-components are provided in section *4.7 Depot*.

3.2 MAD Spatial Topology

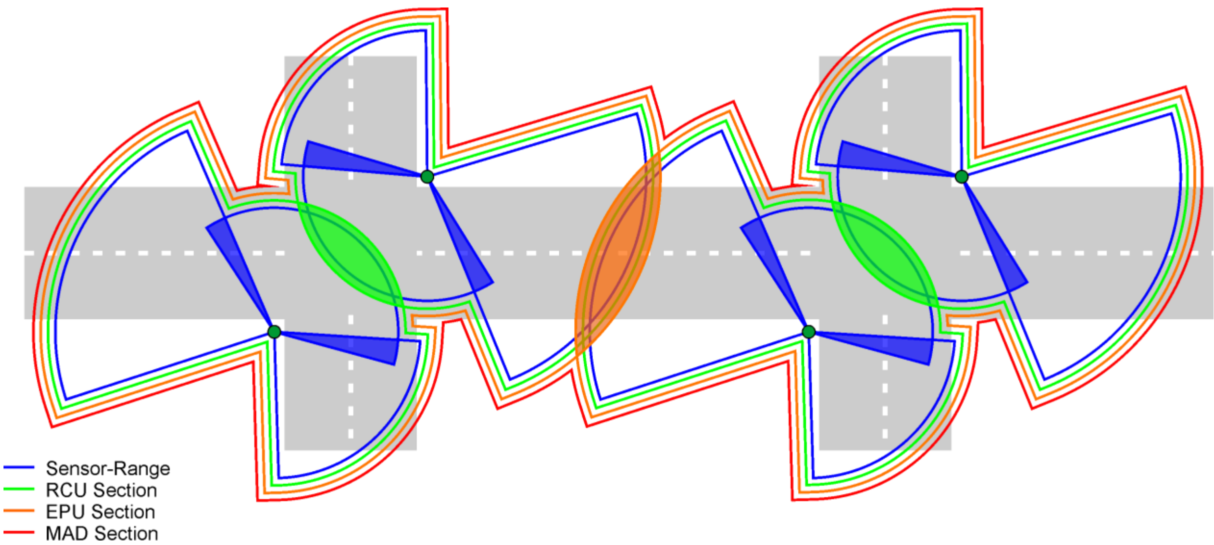


Figure 4: Abstract map view showing the different types of sections and ranges of the MAD components. Overlapping areas of the differ-ent types of sections and ranges are highlighted in the respective color.

For MAD it is important to have a clear understanding of the geometric dimensions and the range and coverage of the different components. As shown in Figure 4, the different components have different ranges, and the ranges are overlapping. This starts already at the sensor level, since different sensors mounted at the infrastructure inside a single RCU will most likely have overlapping sensor ranges, shown in blue. In some cases, it will be an obligation to have overlapping sensor ranges in order to fulfill redundancy requirements for Level 4 vehicle automations, e.g. for MAVs.

Each RCU covers a specific area – the RCU section, shown in green color. Each RCU section is including the ranges of all sensors mounted within this RCU. To ensure a fluid movement of MAVs through an area, several RCUs are required with partially overlapping RCU sections, indicated by the filled green areas in the figure.

As described, a set of RCUs is controlled by an EPU instance, therefore each EPU instance also has a related local operating area – the EPU section, shown in orange. Since an EPU is responsible for controlling all MAVs traveling in the related EPU section, it is important that there is a fluid handover of control at the edge of each EPU section from one EPU to another. This requires that also EPU sections overlap in specific areas, shown by the filled orange area in the figure.

Finally, an MAD provider is responsible for a set of EPUs, integrated in a backend system. Each MAD provider is therefore controlling a specific area covered by all included EPU sections. This area is called the MAD section, which represents the full area where MAD is integrated and therefore possible.

There can be various MAD providers in the future, which would lead to several MAD sections which again may overlap. The related discussion is not part of the current project MAD urban and will therefore not be in scope of this document.

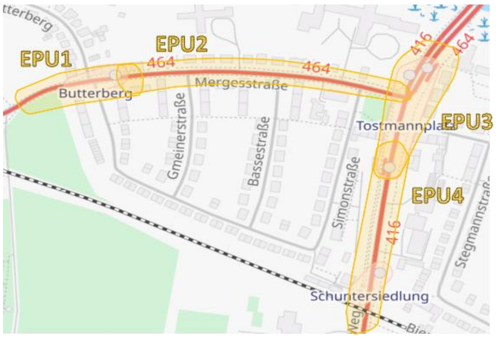


Figure 5: Example map view with a selection of possible EPU sections for a residential area.

3.3 Dynamic automation tasks allocation between vehicle and Infrastructure

One of the goals of the MAD concept is moving as much of the automation Soft- and Hardware to the infrastructure reduces the requirements on the vehicle automation system to fulfilling only fundamental safety functions to reach a safe state in case of a system error (e.g. connection loss to the infrastructure): MAD proposes two different approaches of moving functions to the infra-structure. In "Full MAD", all tasks except drive control are moved to the Infrastructure (Figure 4), making the interface between infrastructure and vehicle a direct driving command (acceleration and steering angle). This approach requires a high frequency V2I connection.

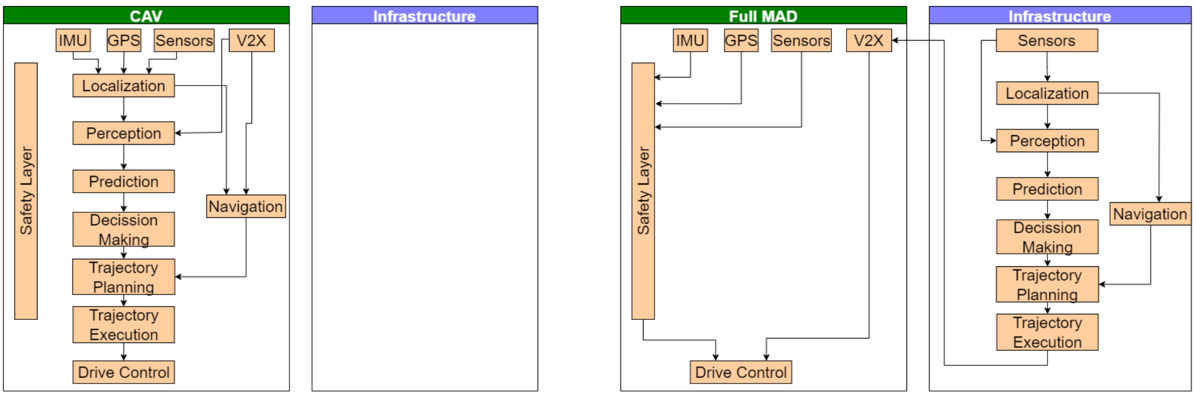


Figure 6 Comparing a typical CAV to Full MAD Architecture

In the second approach "MAD Research" the task of Trajectory Execution remains in the vehicle (Figure 5). Making the interface between V2I Interface a Trajectory calculated by the Infrastructure.

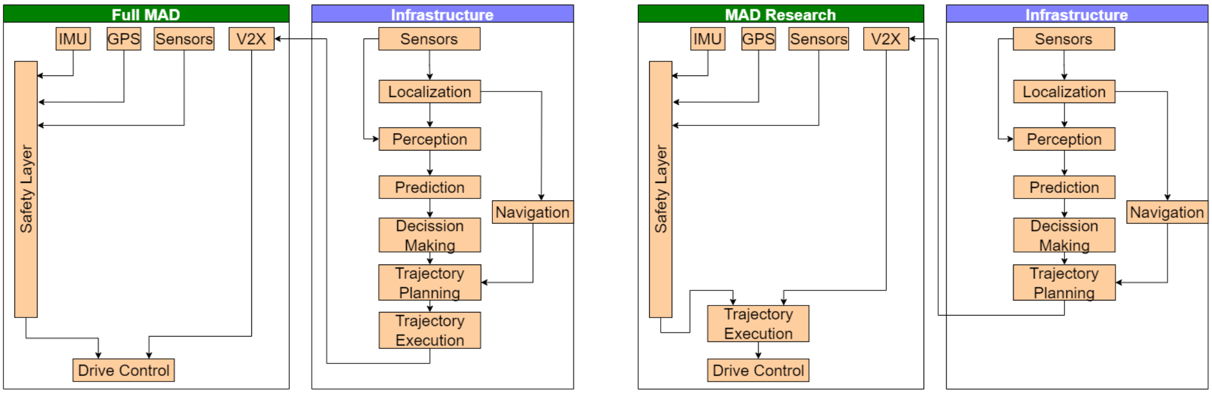


Figure 7 Comparing MAD Research to Full MAD

4 MAD System Components

4.1 Managed Automated Vehicle

Vehicles controlled by the MAD System are called Managed Automated Vehicle (MAV). Since MAVs are controlled by the infrastructure, MAVs can have reduced Automation Soft- and Hardware. However, some sensors and intelligence in the vehicle is still required for safety related tasks like for example performing a minimum risk maneuver to reach a safe state in case of a breakdown of the communication system. MAVs can either be Full MAVs or Guest MAVs. Full MAVs are fully designed to be operated via a MAD system. They don't have a full vehicle-based automation system and can only be operated within areas that are serviced by MAD. Guest MAVs on the other hand have their own fully capable automated driving system and use MAD as a service.

4.2 Road Capturing Unit

Road Capturing Units (RCU) have the task to perceive the traffic scene using sensors like camera, stereo camera, radar, lidar, event camera or others to provide this or object information to the local computation edge. The edge then fuses this information to provide a consistent world view on the uses this locally available information to compute a consistent environment model of the traffic scene, e.g. in an edge section. This includes resolving sensor inconsistencies, sensor uncertainty and noise, registration in a common target world coordinate system and may include estimating additional quantities like velocities, orientation or extent of the traffic objects. The environment model has to be usable by the behavior and motion planning algorithms that use this information as an input.

4.3 Edge Processing Unit

Automation of vehicles in the infrastructure brings the challenge of having high demands regarding low latency as well as high computation power. To meet these requirements, units with high computing power located at the edge are used. These are called Edge Processing Unit (EPU) and aggregate multiple MAVs and RCUs. They offer high-speed connection to the other components of the MAD architecture and are used for tasks like object detection and decision making.

An EPU is a computation unit located at a road segment of interest, e.g. an intersection. It has multiple tasks to perform: real-time object recognition to support the automated driving function, sensor fusion to accumulate the information from all local sensors into a consolidated local environment model, object management in an EPU section and service level data management like making local information available to a global backend infrastructure, e.g. traffic flow information, health/status information, video streams for remote driving.

4.3.1 Information Fusion for Environment Perception in EPUs

The goal of information fusion in an EPU is to construct a consolidated environment model in the roadside infrastructure that contains relevant information about the surroundings that can be used as a basis to derive maneuver decisions and trajectory planning in the highly automated driving function, which may be (partially) moved to the roadside infrastructure. In environment perception, sensor fusion tries to derive, compute and update a model of the static and dynamic objects in the traffic scene with minimal perception errors, e.g. compensate sensor occlusion (no track loss), temporal consistency (no trajectory jumps), spatial consistency (no double detections), no false positive objects, no false negative objects, low positioning, heading, velocity and classification errors. There are three major types of fusion architectures that can be applied in the static MAD infrastructure and that have different opportunities and drawbacks. In the following we give a short overview.

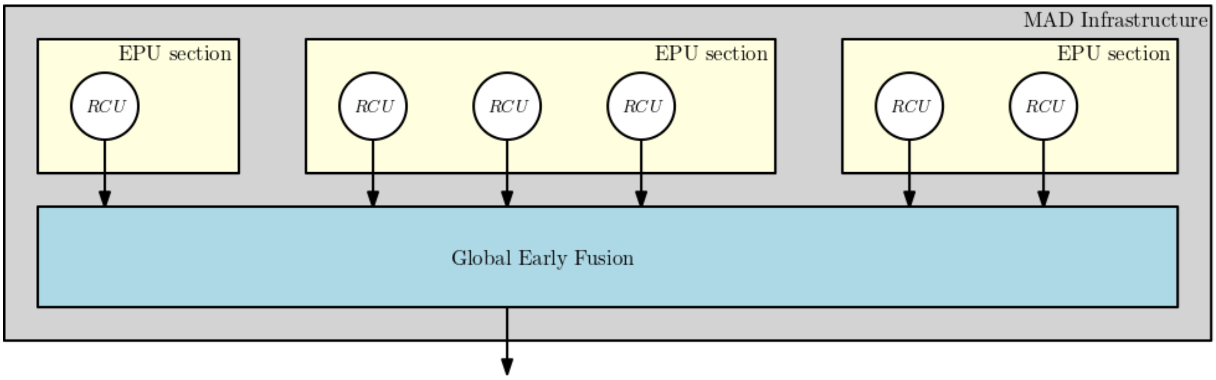


Figure 8: Global Early Fusion. In the global early fusion approach, all raw sensor data is streamed to a global entity and then pro-cessed in a central fashion. This approach typically reaches the highest quality as no information is lost in intermediate steps.

Figure 8 sketches the global early fusion scheme that typically leads to the best fusion and perception quality. All raw sensor data is streamed to a global entity and then computed in a joint fashion. This enables the recognition and fusion algorithm to incorporate all available information which makes the fusion approach the best possible from a data quality perspective. This advantage has major drawbacks related to applicability: commercial sensors may not provide raw data interfaces which makes their integration into an early fusion approach difficult. Especially radar sensors do typically not provide access to the raw data as it is too large and too high frequent to be processed by standard software components. Instead, abstractions like cluster or object lists are the typical interface for this sensor modality. Additionally, very high requirements regarding throughput/bandwidth and latency of the communication medium is introduced as no data is abstracted before sending the data to the global entity. For the managed automated driving use case, global early fusion may need optical fibre connection to each sensor, which might be a challenge in large scale deployment. Anyways, the methods and algorithms for global early fusion may be used in small, spatially related parts, e.g. in EPU sections to provide a locally optimal estimate of the surroundings.

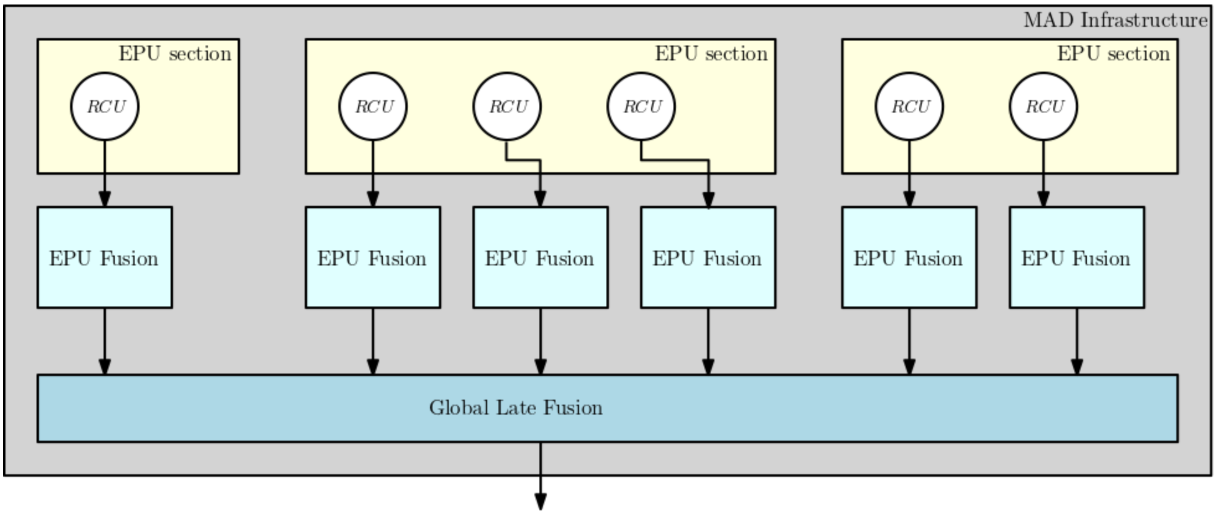


Figure 9: Per Sensor-Fusion with Global Late Fusion. Object recognition is performed per sensor independently and the intermediate data is fused in a global late fusion approach. This topology is very modular but limited in data quality due to information loss.

Figure 9 shows a fusion approach that assumes that every sensor has a computation or precomputation unit that processes data of the sensor before sending it to a global late fusion component. This approach is very modular and scalable as the information is condensed as early as possible in the communication process, so the need for high bandwidth, throughput and low latency becomes less critical than in the Early fusion approach. Additionally, sensor specific computation EPUs may be tailored to the demands of the sensor and do not have to perform general purpose tasks. Anyways, the drawback of this architecture is that information is lost at an early stage in the fusion process, e.g. object positions that are computed per sensor at an EPU may not contain additional information like color or other properties that can be used to support the later fusion process. An additional problem is the so called "Common-Process-Noise"[5] phenomena that occurs in typical Bayesian object fusion approaches. When the preprocessed data is temporarily correlated, different types of fusion methods might be needed than classical Bayesian methods. Another drawback is that typical sensor vendors do not provide preprocessing modules with their sensor, e.g. object recognition. Even if they do, the preprocessing results may have different semantics from other vendors due to the lack of semantically well-defined interfaces in the domain. For instance, one sensor may classify a person riding a bicycle as two objects (a person and a bicycle), whereas a second sensor may classify a single object that is classified as a "bicycle with driver". To consolidate such contrasting information in a global fusion module is a challenging task. Figure 9 shows a fusion approach that assumes that every sensor has a computation or precomputation unit that processes data of the sensor before sending it to a global late fusion component. This approach is very modular and scalable as the information is condensed as early as possible in the communication process, so the need for high bandwidth, throughput and low latency becomes less critical than in the Early fusion approach. Additionally, sensor specific computation EPUs may be tailored to the demands of the sensor and do not have to perform general purpose tasks. Anyways, the drawback of this architecture is that

[5] Campo, Y. B.-S. a. L., 1986. The Effect of the Common Process Noise on the Two-Sensor Fused-Track Covariance,. *IEEE Transactions on Aerospace and Electronic Systems, vol. AES-22, no. 6.*

information is lost at an early stage in the fusion process, e.g. object positions that are computed per sensor at an EPU may not contain additional information like color or other properties that can be used to support the later fusion process. An additional problem is the so called "Common-Process-Noise" phenomena that occurs in typical Bayesian object fusion approaches. When the preprocessed data is temporarily correlated, different types of fusion methods might be needed than classical Bayesian methods. Another drawback is that typical sensor vendors do not provide preprocessing modules with their sensor, e.g. object recognition. Even if they do, the preprocessing results may have different semantics from other vendors due to the lack of semantically well-defined interfaces in the domain. For instance, one sensor may classify a person riding a bicycle as two objects (a person and a bicycle), whereas a second sensor may classify a single object that is classified as a "bicycle with driver". To consolidate such contrasting information in a global fusion module is a challenging task.

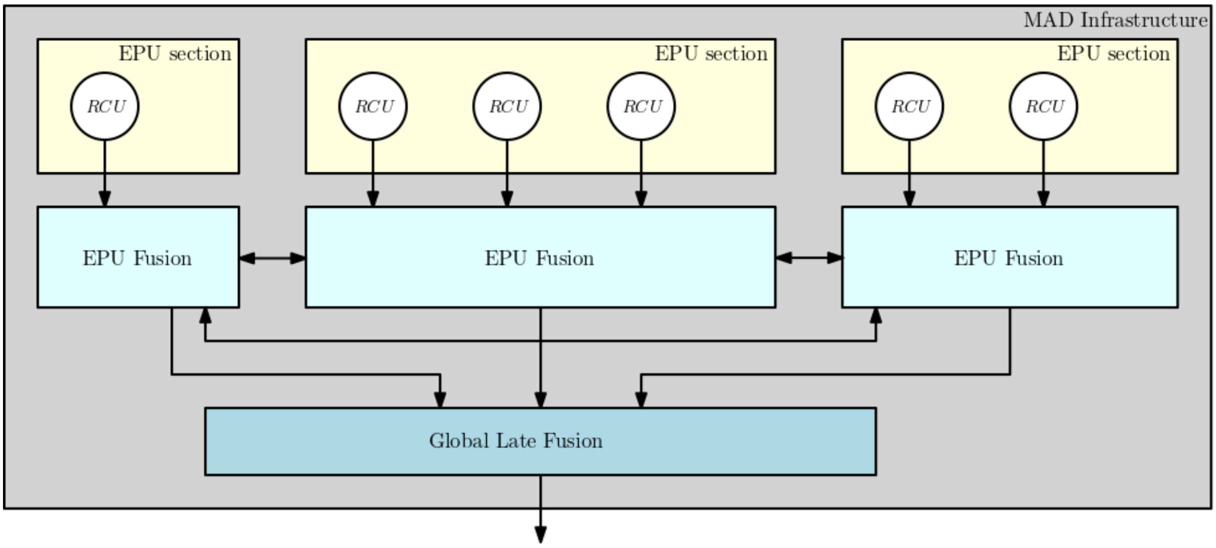


Figure 10: Mixed Fusion approach. Data can be preprocessed and condensed based on the hardware topology. Fusion may happen per Sensor, per EPU and on a global level.

Figure 10 sketches the most reasonable approach to be deployed in real world settings for MAD but comes with a lot of challenges to be explored and investigated.

The approach allows a logical topology where raw information from spatially related RCUs may be fused locally and consolidated in the edge processing unit (EPU). In in extreme case, an extended RCU, e.g. a radar or a camera with included processing unit can directly provide high level information like tracked object positions that then need to be fused on an object level between different EPUs. Environment perception information can be shared between neighboring edge processing units where needed and the global late fusion module is in charge to collect all abstracted data from the EPUs, consolidate it and make it available to the mission and trajectory planning components or other, non-realtime applications. This architecture has multiple challenges. One challenge is to ensure that the flow of information in the system does not contain any cycles that introduce self-fulfilling prophecies in information reasoning. On the other hand, like in Figure 9, preprocessed data that might have temporal correlation has to be fused in the global late fusion module, having altering requirements to the fusion module compared to the early fusion case. Although the most complex architecture, the mixed fusion approach might be the most applicable in reality since it enables the possibility of different module vendors, custom underlying connectivity architectures to which the fusion architecture can adapt to.

All Fusion approaches require that information on spatial and temporal relations between the sensor data is given, meaning that precise timing information for all sensor data (e.g. common high-precision GPS timestamps) and sensor calibration information and the spatial relation is required to be able to make use of the data.

Integration with EPU perception data and in-Vehicle-Sensor data

The integration of perception data into the in-vehicle sensor processing system depends highly on the automation functions and their distribution across the infrastructure and the vehicle.

Option 1 is the joint perception between vehicle and infrastructure. If the vehicle has internal perception sensors and processing, but performs decision making and trajectory planning in the vehicle, a late fusion in the vehicle is required to consolidate the external perception data with the internally perceived perception data.

Option 2 is infrastructure-only perception. In this case, the typically a localization system is needed in the vehicle to transform the received data to a vehicle-relative coordinate system. Since all consolidation of the data happens in the infrastructure, no further information fusion is necessary.

4.3.2 MAV/Vehicle Control in an EPU Section

As previously mentioned, the EPUs play a crucial role in this project. The Edge Processing Units are capable of managing the automated vehicle at three distinct levels: advisory level, planning level, and control level. To effectively carry out these tasks, the EPUs require a comprehensive understanding of the dynamics of the road. Considering the stochastic nature of the environment, accurately predicting the intentions of each road user is essential.

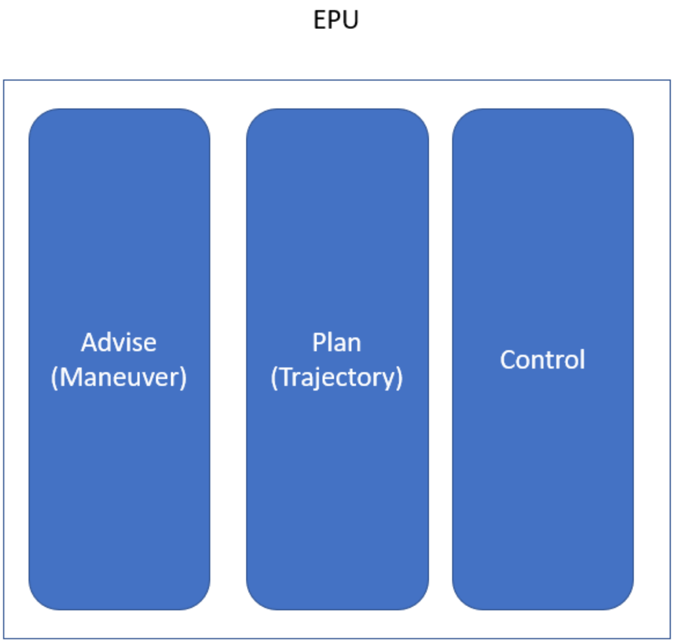


Figure 11: EPU levels: advisory, planning and control

In this context, it becomes necessary to have intention prediction models with the capability to forecast how the traffic scenario will evolve based on the physical state of vehicles, potential maneuvers, and interactions among traffic participants within the upcoming seconds. To predict the behavior of each road user, a set of representative trajectories is defined for each vehicle, taking into consideration the map information. These trajectories are based on a simple kinematic model to ensure realism and represent specific maneuvers and behaviors that the vehicle could potentially adopt in the future.

The approach considers in a game theoretic framework the interactions and reciprocal influence between the vehicles. The application of the Bayes' theorem allows to correct the rational prior estimate with the data coming from the road, by taking into account the current observable maneuver.

.

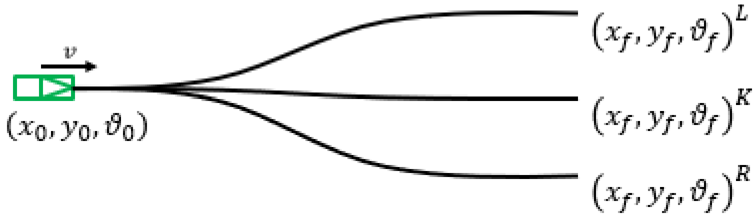


Figure 12: set of representative trajectories

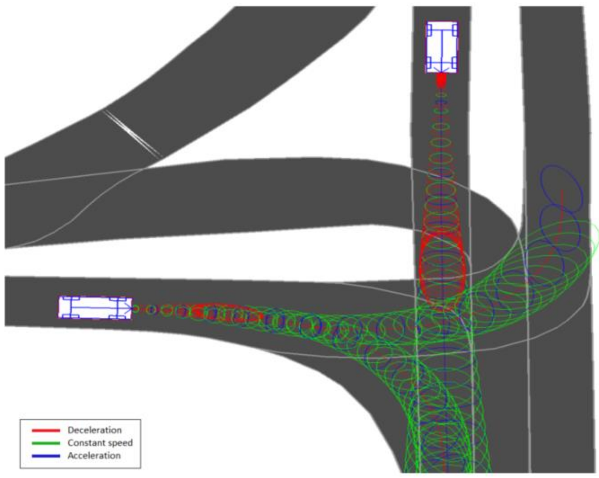


Figure 13: Example of trajectories

Since the output consists of a set of trajectories with varying probabilities, the trajectory with the highest probability is the one to be considered. The module dealing with trajectories allows for different levels of information extraction.

At the advisory level, the EPU can provide specific suggestions to a vehicle based on the intentions of other road users. These suggestions may include maneuvers such as maintaining the current lane or changing lanes at a particular position and time. The vehicle's decision-making module can take this advice into consideration and either accept or decline it.

At the planning level, the EPUs can transmit the complete trajectory to the vehicle. This trajectory encompasses a vector of information that includes the vehicle's position, velocity, acceleration, and other relevant parameters at different time intervals. By receiving the full trajectory, the vehicle gains a comprehensive understanding of the planned path. Upon receiving the trajectory, the vehicle's decision-making module can compare it with the trajectory generated by its own trajectory planner. This allows for a thorough analysis of the proposed path, taking into account factors such as efficiency, safety, and adherence to traffic rules. The decision-making module can then determine whether to accept the provided trajectory or reject it in favor of its own planned trajectory.

At the control level, the vehicle can be governed by the EPU. This typically occurs in situations where there is a failure in vehicle automation or perception systems. In such cases, the EPU takes over control of the vehicle. By leveraging the available information about the vehicle, such as its size, relevant control parameters like steering angle and longitudinal acceleration can be derived from the trajectory. This information is then transmitted to the vehicle at a high frequency. The EPU's controller utilizes this information to guide the vehicle along the planned trajectory. Through this control mechanism, the vehicle can be effectively steered and managed by the EPU, ensuring its alignment with the intended path and facilitating a safe and efficient navigation process.

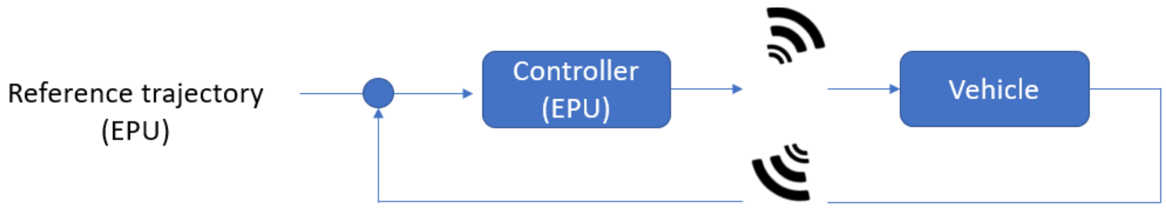
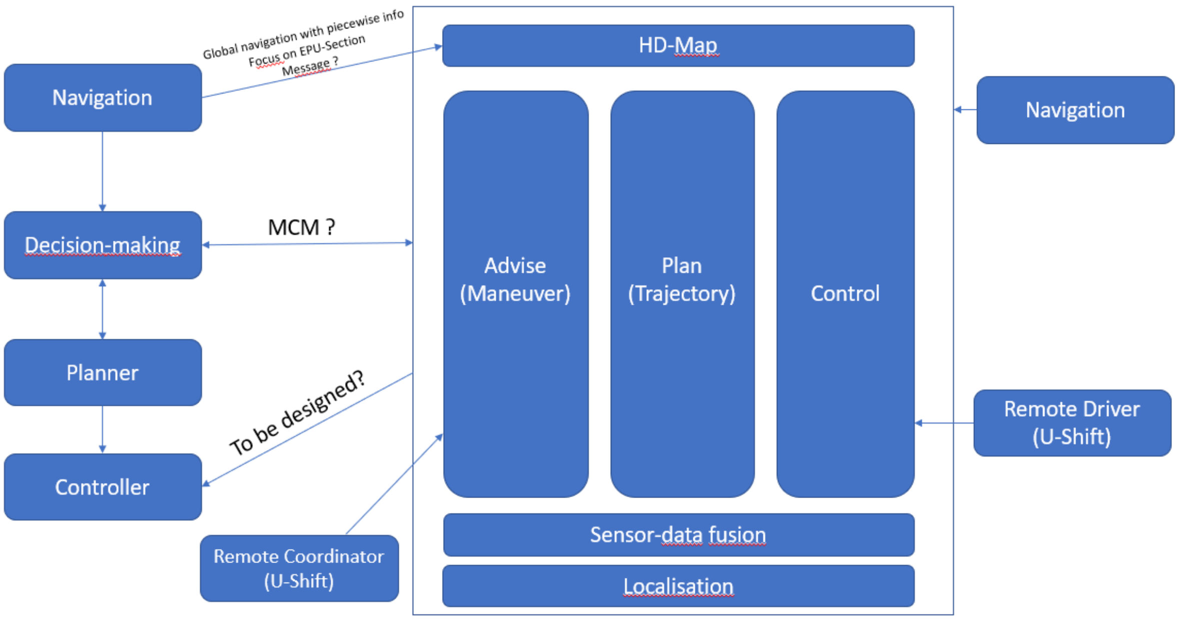


Figure 14: Control level

Module Interconnectivity:

The edge processing unit (EPU) relies on information from other modules to ensure proper functionality. In this section, we will delve into the interfaces and information flow between the EPU, other modules, and the vehicle.

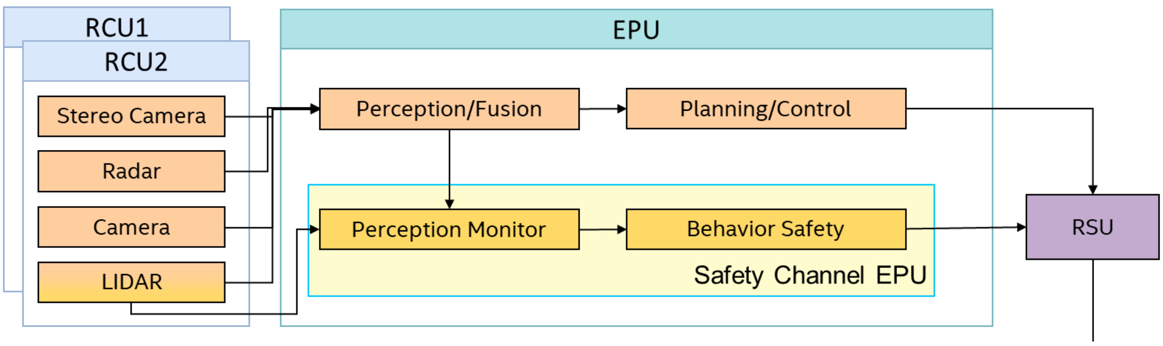
In the figure below, we present the block diagram illustrating the interface between the Edge Processing Unit (EPU), the vehicle, and other modules. In this representation, the vehicle is depicted solely in terms of its automation functionality, encompassing navigation, decision making, planner, and controller modules.



4.3.3 Data Management in EPU Sections

To enable communication between RCUs, the backend system must be designed in a decentralized manner. For this purpose, a message server is deployed in each EPU section. This provides the advantage that communication between the RCUs can take place with minimal latency. Furthermore, messages relevant to the backend can be pre-processed, thereby keeping the necessary bandwidth to the backend system low. This allows operators/users of the backend system to access important information without overburdening the system. Additionally, a decentralized architecture offers a direct connection to the vehicles and video streams for remote control.

4.3.4 Safety Layer in EPU Sections



Safety always has to be ensured on the system level. Therefore, every component of the MAD system has to be considered. To ensure safety when performing perception and planning tasks predominantly in the EPU, it is of utmost importance to establish safety measures to prevent traffic participants from harm. Considering Full MAD with MAVs equipped with a reduced set of on-board sensors, a distributed safety approach with safety components running at EPU level to safeguard the perceived environment model as well as the multi-agent planning stage becomes mandatory. To cover communication and other remaining risks at vehicle level, the EPU safety components have to be combined with fallback safety components running on-board of the vehicle. A possible implementation of safety components deployed in the EPU is briefly sketched in the following, whereas a more detailed description of the overall safety concept and system architecture can be found in Sec. 5.1.

The safety channel envisioned is implementing a monitoring approach where (simpler) monitoring functions check and correct the results of complex, potentially AI-based, primary perception and planning components running in the EPU. The safety approach consists of two core components: the perception monitor and the behavior safety module. These provide safe fallback strategies in case of failures or errors in the primary perception or planning modules. The perception monitor runs in parallel to the primary perception algorithms and uses dependable 3D depth data, typically from 3D LIDAR sensors, as input. By deploying classical, i.e., non-AI-based algorithms, the perception monitor validates the dynamic objects detected by the primary perception channel against the depth data. In case of missed detections, the monitor adds these missing objects and passes the revised object list to the behavior safety component. The behavior safety component runs in parallel to the primary planning algorithms and constantly monitors the safety envelope of all traffic participants based on the revised object-list received from the perception monitor. In case the safety envelope is compromised, the behavior safety module aims to bring the respective vehicle back to a safe state such that it will not cause a collision if other road users behave reasonably foreseeable.

4.4 Backend

The basis of the backend is the capture, provision, and processing of data from all road users and traffic infrastructure in digital twins, as well as the control of automated vehicles. Therefor the backend must link, preprocess, analyze, and monitor various data sources. To accomplish this, multiple services/components are required to handle these tasks. Figure 15 shows an overview of the backend services.

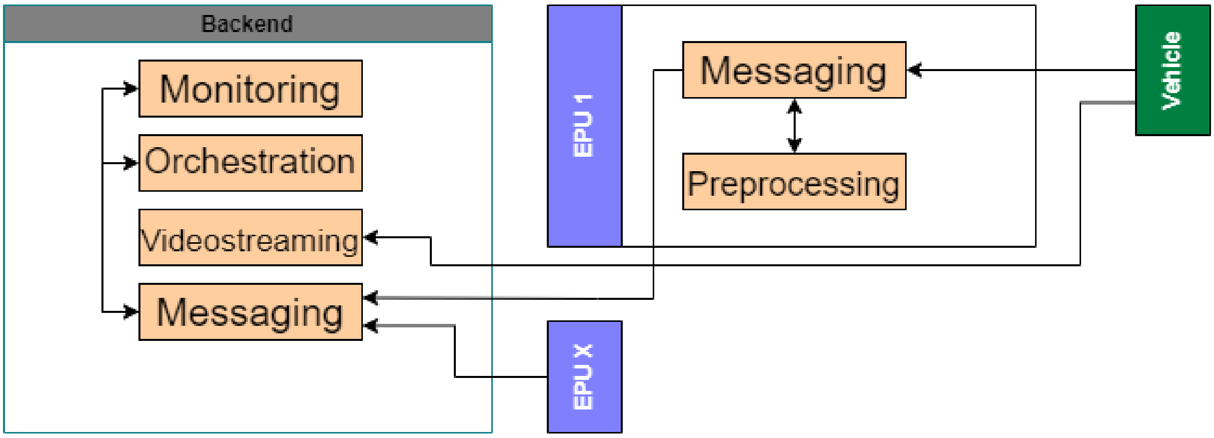


Figure 15 Overview of the backend services and the communication flow.

4.4.1 Backend Services

The backend system consists of multiple services that are essential for fulfilling tasks within the MAD system. The following sections provide a detailed description of the fundamental backend services.

Messaging Service

The Messaging Service acts as the backbone for all data communication within the MAD system. It ensures reliable and low-latency data exchange between the backend, edge servers, and

MAVs, using protocols such as MQTT to facilitate message delivery. At the EPU level, the service collects data from various MAD components, including MAVs, road sensors, and other infrastructure, performing initial pre-processing to reduce data volumes. The backend then aggregates this data from multiple EPUs and integrates it with other system inputs for deeper analysis. Additionally, the Messaging Service handles both real-time communication for critical tasks and asynchronous messaging for less time-sensitive operations, such as data logging or system diagnostics. This design ensures that essential system processes remain responsive while maintaining scalability for long-term data management needs.

Videostreaming Service

The Videostreaming Service processes video feeds from MAVs and infrastructure-mounted cameras, providing a real-time view of roadway conditions. MAVs send their video streams directly to the backend, where they are exposed to the MAD Operation Center (MOC). Operators in the MOC rely on these high-resolution streams to make informed decisions during teleoperation or to address system anomalies in real time. The service ensures that video streams are transmitted with minimal latency and supports compression techniques to optimize bandwidth usage.

Monitoring Service

To ensure the overall system's safety, it is necessary to monitor all road participants and infrastructure components. These may include, for example, RCU's, MAV's, EPU's, or other vehicles. This allows for the early detection and prevention of system errors. To efficiently achieve this, digital twins must first be created for all components. Subsequently, received data can be associated with, processed, and evaluated against the digital twin. This enables rapid decisions to be made regarding the status of individual participants. A web interface in the MAD Operation Center serves as the visualization platform.

Orchestration Service for the teleoperation of the MAV's

Another important task of the backend is the provisioning of an orchestration service for the teleoperation of the MAV fleet. Since the MAD Operation Center is a multi-user system, it is necessary to have an orchestration service to ensure that only one user at a time has access to a single teleoperated MAV. Another task of the service is establishing a connection to the vehicles, ensuring that the correct control commands and video signals are available to the respective users.

4.5 Customer API

The customer API is the general interface of the MAD components to the outside world. It is allowing access to parts of the backend components, allowing to include/import external data and to provide/export internal data.

The customer API therefore is the generic interface to all additional services. It can be used e.g. to link

- To a fleet management system
- To a logistics system
- To public transport operators
- To other infrastructure components such as traffic light controls
- To weather, disaster, event and congestion data
- To road works and maintenance systems
- To monitoring tools or systems, dashboards
- To energy management systems

- Etc.

Since this requires several different software interfaces, this customer API will not be defined in detail in the project. Nevertheless, it is currently identified that the customer API may use additional upcoming technologies, e.g. based on the Eclipse Dataspace Components (EDC)[6], Gaia-X[7], or the Mobility Data Space[8], OPC-UA, REST-API, MQTT.

4.6 MAD Operation Center

The MAD Operation Center (MOC) is to be set up within the facility where, e.g., municipalities or related contractors control the local traffic, particularly public transport. An example for such a facility is a Traffic Management Center (TMC) or the control center of the local public transport provider.

The MOC includes different workstations for the different tasks that have to be done. Multiple Remote Operation Centers on different locations are possible. The role of the MOC in the MAD ecosystem is similar to traffic management centers for buses and rail transport systems.

4.6.1 MAD Coordinator Workstation

The key workstation located in the MOC is the Remote Coordinator Workstation, the workplace of the **Remote Coordinator (RC)**. The RC is responsible for smooth operations in the MOC and is in charge of the other employees in the MOC, i.e., MAD System Operator (MSO). Remote Driving Operator (RDO), and Service Technician (ST) (role descriptions see below). The RC may fulfil the role of the Technical Supervisor ("Technische Aufsicht") as specified by German law (StVG/AFGBV). The RC is responsible for ensuring that the legal framework conditions or specifications are complied with. The RC closely monitors MOC operations overall, is in charge of resolving disturbances of any kind and allocates tasks to the other roles in the ROC in order to achieve this. The RC accepts reports on incidents, including emergency calls, coordinates operations when irregularities occur and defines tasks to be fulfilled to overcome those incidents. In case a higher-level maneuver is to be conducted, the RC provides remote assistance, e.g., by setting waypoints to guide towards a feasible trajectory. If the required task does not fall under remote assistance, the RC continuously exchanges information on the current situation with MSO, RDO, ST, and further roles, assigns tasks to them, monitors the tasks' execution and keeps track of their progress. Furthermore, the RC communicates with actors both internally and externally to overcome the irregularities, such as dispatchers, information managers, and passenger assistants, but also police, firefighters, and medical emergency services. Specifically, the RC is required to inform the passengers once a minimum risk maneuver has been executed by the MAV. In addition, the RC files reports on irregularities that include a detailed description of the event, measures taken, and staff involved. Nevertheless, the RC is not a dispatcher.

4.6.2 MAD System Workstation

The MAD System Workstation is the workplace of the **MAD System Operator (MSO)**. His roles are (1) configuring, monitoring, evaluating, and updating the automation *software* both for all the components and specific parts of the area of operation, including reanalyzing video images and post-labeling and categorization of not yet classified objects, (2) maintaining *system components* and monitoring failures thereof, (3) making sure the operation can continue by supporting the Remote Driver in analyzing *failures*, and (4) creating *logfiles* of any incident.

[6] https://github.com/eclipse-edc/
[7] https://gaia-x.eu/
[8] https://mobility-dataspace.eu/

4.6.3 Remote Driving Workstation

Remote driving is a relevant part of MAD. So, another essential workstation located in the MOC is the Remote Driving Workstation, the workplace of the **Remote Driving Operator (RDO)**. It is designed to execute (1) manual driving operations at a single driveboard, such as steering, accelerating, and braking, using the direction indicator, and shifting the gears (forward/backward), as well as (2) docking on and off capsules. If a given capsule is equipped with doors, the RDO will further be responsible of (3) opening and closing the capsule's doors.

4.6.4 Further workplaces

Maintaining the operation of a fleet of MAVs will require several other workstations and staff members in the MOC, some of which are also required by law (StVG/AFGBV). They will be further specified in later stages of the architecture. Here is a summary of possible tasks for the additional staff members: Dispatching/Scheduling, fleet management, energy management or customer/passenger support.

4.7 Depot

The Depot is another physical location in the MAD architecture. It provides a central hub for storing unused MAVs and maintaining and repairing them if necessary.

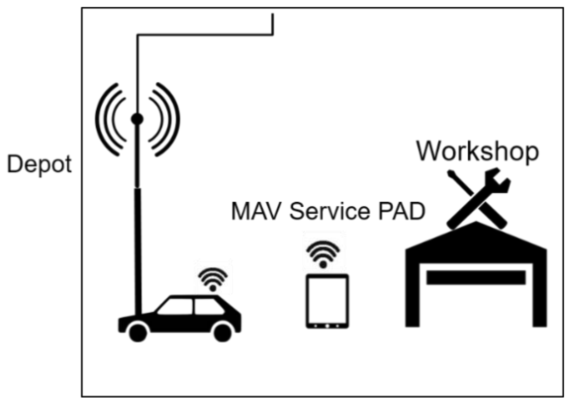


Figure 16: Depot

4.7.1 Workshop

There is a workshop located in the depot where **Service Technician (ST)** can do all the required maintenance and repairs of the MAVs. If MAVs are scheduled for maintenance or have issues beyond the scope of the MSO, they are sent to the workshop. There the service operator can use a MAV Service Pad to take control of the MAV and maneuver it inside the Depot and workshop.

4.8 MAD Communication

In general, communication plays a significant role in MAD Urban, since all components require stable and often redundant communication channels between each other. Esp. the control link between the vehicles (and esp. MAVs) and the EPUs needs to be trustworthy, stable, safe, and with very low-latency. For backward compatibility and for including existing and upcoming C-ITS messages, it is also important to be compatible to current C-ITS, ITS-G5 and ETSI standards.

In case of remote operation, there is also the necessity to have low-latency connections between the remote operator and the vehicles. The remote operator also needs to have an excellent overview on the local situations, which requires a detailed and multi-perspective video transmission from the vehicles and the local RCU sensors. This adds also adds high bandwidth to the requirements.

Since the performance of the whole system is important and needs to be monitored, and since also decisions, states and sensor data needs to be logged to allow e.g. forensic investigations in case of incidents, further connectivity requirements esp. between ECUs and the backend system need to be added.

MAD Urban is not investigating the detailed capabilities of different networks and protocols, but is collecting connectivity requirements and also implements first connectivity solutions. The detailed protocols are described in section 5, in relation to the functional description.

5 System functional architecture

Figure 17 shows an overview over the functional components of the MAD automated driving function. The figure contains three vehicles with different capabilities and sensors. The MAD concept allows that all types of vehicles can profit from the MAD infrastructure in an MAD section.

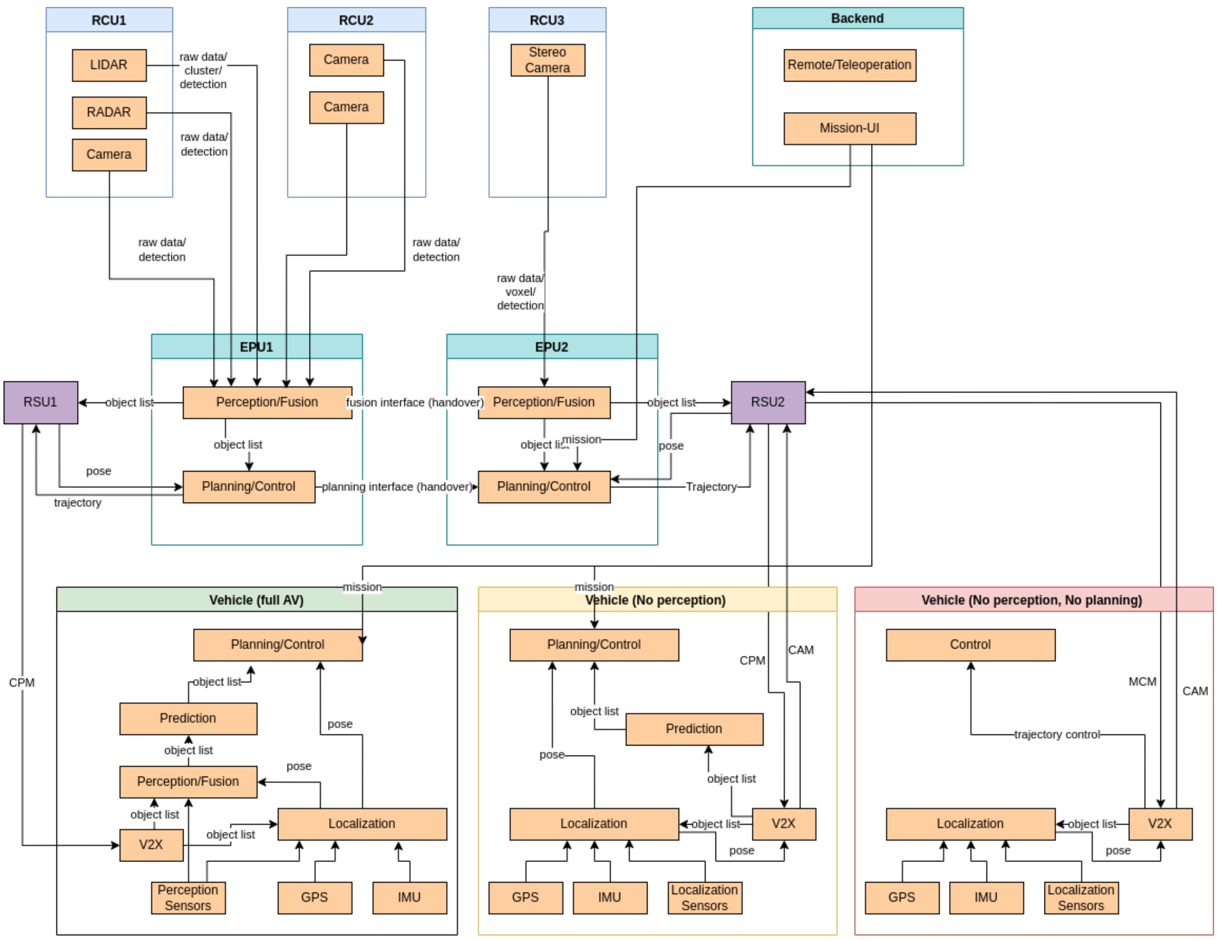


Figure 17: Overview of the Automated Driving Function Architecture (for better overview without the Safety-Modules).

5.1 Vehicle-Localization

Vehicle localization is a crucial functionality of the driving function, regardless of the allocation in infrastructure or within the automated vehicle. The vehicle localization module provides a high precision, high frequency object position and orientation. In general, this can be given in different coordinate systems: global coordinates on a reference geoid or ellipsoid (e.g. latitude, longitude in WGS84), a local tangential Cartesian coordinate system, e.g. East-North-Up (ENU) on a surface reference position or grid-based local coordinate systems like Universal Transversal Mercator coordinate system (UTM). Typical vehicle localization systems make use of the Global Positioning System (GPS) or similar services like GLONASS (Russian), GALILEO (European), BEIDOU (Chinese) and combine this global position measurements with local odometry data, e.g. from wheel odometers and or inertial measurement units (IMU). Depending on the vendor, the localization data then is typically given in global coordinates. Relative positioning approaches include Simultaneous Localization and Mapping (SLAM) methods that may use visual sensors (Visual-SLAM), LiDAR sensors (LiDAR-SLAM) or even Radar sensors. These approaches build locally registered localization maps from mapping drives in the operational design domain of the driving function and then later use the created map to match it with current sensor data which leads to a local positioning relative to the recorded map. Here, localization information is typically given in local Cartesian coordinates and no global navigation may be needed.

In the MAD concept, it is possible to use external positioning measurements from the perception system of the roadside infrastructure and incorporate it in the localization process of the vehicle with the aim to gain robustness against common localization errors like GPS drift in urban canyons and other occurring global positioning disturbances in global positioning.

5.2 Perception, Object Detection and Tracking

5.2.1 Coordinate Systems

- Local Cartesian vs Sensor coordinates vs CPM coordinates

5.2.2 Interfaces

- Table with Interfaces (see lines fig Overview)
    - Object Lists
    - CPM
    - Handover interfaces
    - Raw interfaces
    - Localization interfaces

5.3 Automated Driving

Automated driving functionality, also referred to as autonomous driving or self-driving technology, encompasses the ability of a vehicle to operate and navigate without direct human intervention. This capacity relies on the integration of diverse sensors, computing systems, and algorithms to perceive the surrounding environment, make informed decisions, and effectively control the vehicle's movement.

The concept of automated driving functionality is commonly categorized into distinct levels or stages of automation, as defined by the Society of Automotive Engineers (SAE) International. These levels range from Level 0 (indicating no automation) to Level 5 (representing full automation), representing varying degrees of control and intervention by the automated system.

At the core of this functionality lies "vehicle automation," which consists of several modules, each serving a specific purpose. Let's explore them in detail:

- Navigation Module: This module finds a route which connects the current position to the destination using the information of a digital map.
- Perception Module: This module employs a range of sensors such as cameras, lidar, radar, and ultrasonic sensors to collect comprehensive data about the vehicle's surroundings. These sensors aid in object identification, road markings detection, and determination of the vehicle's position in the environment.
- Decision-Making Module: Equipped with sophisticated algorithms and computing systems, the decision-making module analyzes the information gathered by the perception module. It evaluates potential routes, predicts the behavior of other vehicles and pedestrians, and makes decisions concerning the vehicle's trajectory and actions.
- Trajectory Planning Module: Building upon the decisions made by the decision-making module, the trajectory planner computes an optimal trajectory for the vehicle, considering vehicle dynamics, road conditions, and traffic scenarios.
- Control Module: The control module ensures the vehicle adheres to the planned trajectory by precisely manipulating the vehicle's acceleration, braking, and steering systems. It guarantees that the vehicle moves in a safe and efficient manner, following the intended path.

By combining these modules, automated driving functionality empowers a vehicle to autonomously operate and navigate through diverse traffic scenarios, promoting heightened levels of safety, efficiency, and convenience.

5.4 Remote Operation

For situations where the MAV reaches its operational limitations, e.g. regarding its ODD, the backend offers a web application for the Remote Coordinator in the MAD Operation Center to provide technical assistance to the MAV. There is no direct communication between MOC and MAV or RSU.

- Sitting on top of the services provides by the backend, the Remote Operation Interface offers the following functionality:
- Task Manager: Events of MAVs requiring assistance are shown in the Task Manager as tasks corresponding to the MAV’s request. A Remote Coordinator can accept these tasks to work on their resolution. The required exchange of information uses the Messaging Service of the Backend.
- Video Streaming: In order to support the RC in assessing the situation, suitable video streams from the vehicle and the infrastructure can be displayed via the video streaming service of the backend. WebRTC is used as the protocol for low-latency transmission.
- MAV State Presentation: The vehicle status, e.g. position, schedule, technical status, of the MAV supported in the active task is visualized in an overview and a map view based on an external map service.
- Maneuver advisory: To assist a MAV, specific messages can be sent, e.g. to guide the trajectory planning.
- Voice Communication: The interface allows voice communication with vehicle passengers and authorities.

5.5 Responsibility

Work in progress

5.6 Handover

To operate a vehicle with the MAD concept a continuous connection to an Edge Server, the backend and the Remote Center are required. To achieve this a there have to be Handover protocols for entering the next subsections as well as sections.

Since the RCUs are connected to the same Edge-Server, a handover between Sections can work similar to a cellular network or mesh W-LAN. The maximum time this handover is allow to take is specified in Link as 10 ms.

The handover between sections is a bit more complicated, because the automation and message handling has to be handed over between Edge Servers. The advantage is that the routes of the vehicles are following are known by the system. Because of that a handover between Sections can be planned in advance. To guarantee a safe and continuous connection a maximum handover time of 10 ms should be achieved.

5.7 Safety

In MAD, a considerable part of the automation stack will be executed at edge-/infrastructure-level. Sensors are installed at fixed locations/sensor poles and connected to Edge Processing Units or the Mobile Edge Cloud of the network service provider. The primary perception, planning and decision-making functions are supposed to run in the Edge/Cloud. In this configuration the automated vehicle only requires a minimal set of sensors and needs to be able to receive a planned trajectory from the edge/infrastructure and to follow the given trajectory.

In a configuration with the automated driving functionality being performed predominantly in the edge/infrastructure, establishing safety measures to prevent other traffic participants from harm is even more important than in a solely vehicle-centric setup. The primary perception as well as the primary planning functions of a complex automated driving system typically rely on AI methods which makes validation and certification of such systems with respect to automotive safety standards such as ISO 26262 (FuSa) or ISO 21488 (SoTIF) difficult. While validation of the MAD system in simulation, on test tracks and on road tests is required, an online monitoring architecture is applied using a significantly less complex monitor that checks the correctness of the output of the complex function blocks of the primary functions (see Figure 18). This monitor catches safety relevant errors of the primary function and eliminates potential negative impact of such errors on downstream applications (see Figure 19). At the same time, the monitoring function should not decrease the availability of the primary system exceedingly to not limit the intended functionality of the overall system to be monitored.

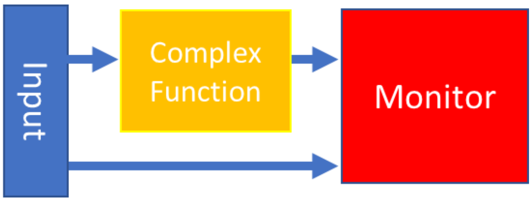


Figure 18 Monitoring architecture, ISO, "26262: Road vehicles-Functional safety," International Standard ISO/FDIS, vol. 26262, 2011.

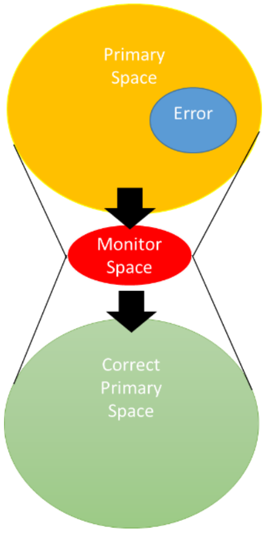


Figure 19 Safety critical errors in the primary space are corrected by the monitor.

The proposed safety monitoring functionality for MAD is split into a safety channel executed within the EPU and an adjacent safety channel executed locally on each of the MAVs (see Figure 20).

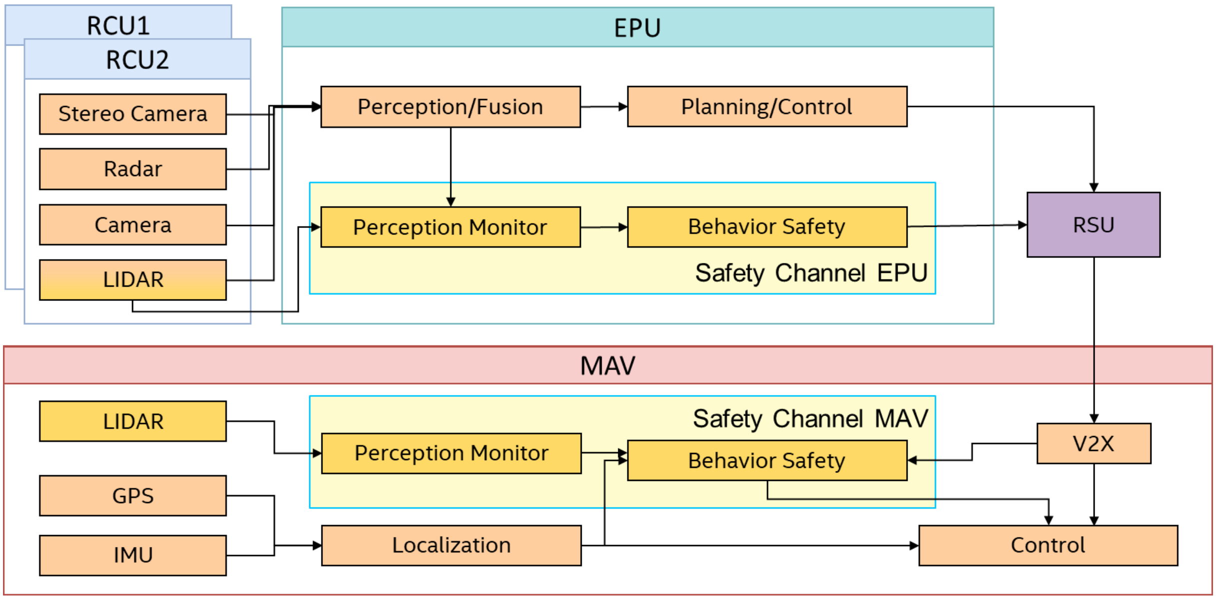


Figure 20 The proposed distributed safety layer to protect from errors within the primary perception, fusion, planning and control functions executed within the EPU to safeguard the MAV.

5.7.1 Safety Channel EPU: Edge Perception Monitor

Parallel to the primary perception algorithms, a perception monitor operating on dependable 3D depth data input is executed in a safety layer within the EPU. The monitor deploys classical, that is, non-AI-based algorithms to validate the list of dynamic objects detected by the primary perception module against the depth data. In case of missed detections in the primary object list, the monitor updates it by adding those missing objects before passing the revised object list to the behavior safety component of the edge safety channel. In contrast to a perception monitor directly within a vehicle, the deployment based on the infrastructure sensors raise additional challenges. For instance, blind spots due to occlusions from static and dynamic objects in the sensors' field of view could occur. Furthermore, the safety relevance of dynamic objects is not necessarily corelating to their distance from the infrastructure sensors. Therefore, the placement and field of view of the sensors within the infrastructure setup is directly connected to the system safety.

5.7.2 Safety Channel EPU: Edge Behavior Safety

Parallel to the primary planning algorithms, a behavior safety component is executed in the edge safety layer. To determine if the ego vehicle is in a safe state, the behavior safety component performs a continuous check of the ego vehicle's safety envelope and defines a proper response in case the safety envelope gets compromised. This response action varies depending on how the safety envelope was compromised and aims at bringing the ego vehicle back to a safe state such that it will not be the cause of a collision if other road users behave reasonably foreseeable. This safety check is performed for all MAVs and the check results are reported to the respective safety channel of the MAV using V2I.

5.7.3 Safety Channel MAV: Vehicle Perception Monitor

The MAV receives the safety check results from the Edge Behavior Safety component. Because of the remote nature of the safety results, these cannot be applied directly to the MAV without consideration of additional aspects. Compared to an in-vehicle deployment, the latency of the transmission of the remote safety check results varies over time, so that an additional cross check of the local environment with the considered objects states on the safety check is performed to ensure the provided safety check results are still valid. For this check a vehicle local LIDAR or comparable dependable depth sensor source is deployed. In case of major discrepancies, the vehicle's minimal risk maneuver is triggered.

5.7.4 Safety Channel MAV: Vehicle Behavior Monitor

It is very likely that the vehicle's view using local sensors deviate from what the infrastructure has detected. Especially if one considers the latency effects introduced by the V2I communication. Therefore, there should be the possibility to perform basic local checks for the vehicle's safety envelope in respect to all locally observed obstacles that might have not been considered by the EPU safety layer yet. Like this, minor discrepancies to the EPU safety layer can be safely mitigated to increase the availability of the overall system.

5.8 General Automation Functions

Work in Progress

5.9 MAD specific Functions

Work in Progress

5.10 MAD Start-Up

Start-Up Sequences for the whole MAD system, Work in Progress

6 Interfaces and protocols

Work in progress

7 Example implementations

7.1 Testbed Lower Saxony Integration

Work in progress

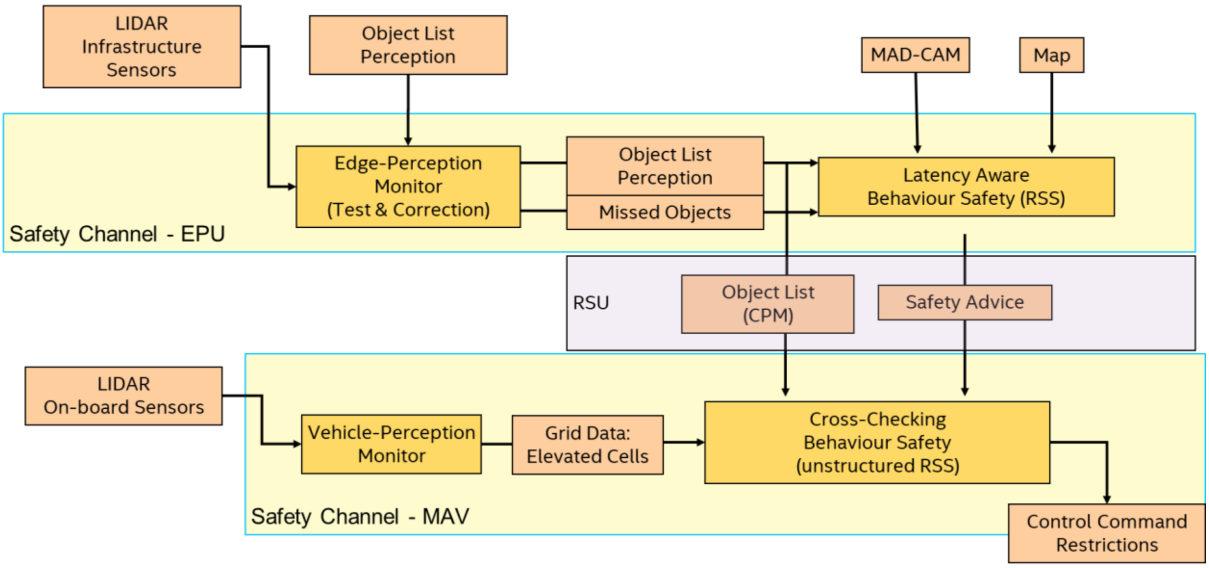
7.2 TAF-BW Integration

Work in progress

7.3 U-Shift Capsule exchange

Work in progress

7.4 Distributed Safety Layer



In the following, a more detailed description of the actual implementation of the sketched distributed safety architecture is provided.

The Edge Perception Monitor in the EPU receives the LIDAR data of all RCUs and distinguishes static from dynamic grid cells in an occupancy grid[9]. The detected dynamic grid cells are matched against the object list from the primary perception functionality of the EPU. Dynamic grid cells which cannot be explained by detected objects are passed in form of (virtual) missed objects (without classification) expanding/correcting the object list.

In addition to the corrected object list, the Edge Behavior Safety Component of the EPU receives information on the MAVs via the V2X CAM messages as well as high-definition map data and applies safety checks based on Responsibility-Sensitive Safety (RSS), a mathematical safety model introduced by Mobileye on all MAVs[10]. The safety checks are performed applying multiple configurations considering different potentially expected latency values (i.e. multiple possible reaction times of the MAV are considered on the evaluation of the respective safety envelope). The resulting safety advices are sent out via V2I to the MAVs.

Within the MAV a local vehicle perception monitor is deploying the onboard LIDAR sensors to calculate all cells of an occupancy grid that are considered to be elevated from the ground[11]. This is provided as input to the in-vehicle cross checking behavior safety component.

The behavior safety component in the vehicle receives the safety advice from the EPU and selects the results matching best to the actually observed latency values. Furthermore, it cross checks

[9] C. Buerkle, F. Oboril, O. Zayed and K. -U. Scholl, "HistoGrid: Robust LiDAR-Based Traffic Monitoring," *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*, Bilbao, Spain, 2023, pp. 209-214, doi: 10.1109/ITSC57777.2023.10422158.

[10] B. Gassmann, F. Oboril, I. Alvarez and K. Scholl, "An Online Safety Guard For Intelligent Transportation Systems," *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, Indianapolis, IN, USA, 2021, pp. 2575-2581, doi: 10.1109/ITSC48978.2021.9565101.

[11] C. Buerkle, F. Oboril, J. Burr and K. -U. Scholl, "Safe Perception - A Hierarchical Monitor Approach," *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, Macau, China, 2022, pp. 71-78, doi: 10.1109/ITSC55140.2022.9921793.

the object list of the EPU in respect to the local environment provided by the elevated grid cells of the vehicle perception monitor. All local observations that have not been considered by the EPU are fed into a local safety check using RSS calculations for unstructured environments (not considering structured lanes) to ensure safety in respect to these. The local RSS check results are then combined with the EPU check results and forwarded in form of control command restrictions to the vehicle controller.

8 Discussion / Next steps

Work in progress

9 Annex - RCU Components

Work in progress

9.1 Road Side Sensors

Road Side Sensors can have different measurement principles, modalities, advantages and disadvantages. To support automated driving functions, the major goal of roadside sensors is to make it possible to detect, classify, track and predict relevant participants of a traffic scene. Especially vulnerable road users (VRUs) (pedestrians, bicycles, scooter drivers, etc.) and highly dynamic or critical participants like driving cars, vans, trucks or trains are of major interest to ensure safe automated driving.

The following section gives a short overview over the sensor principles/processing algorithms and advantages/disadvantages of typical/possible roadside sensors.

9.1.1 Road Side Monocular Camera (RSMC)

Monocular cameras are passive sensors that capture the visual spectrum of a traffic scene and provide color information (e.g. RGB-channel) about the scene. They typically provide an image that follows a certain encoding and that has to be postprocessed to gain information about object information in the image (locations of persons, classification of object, etc.).

Major advantages of monocular camera sensors are their ability to capture very dense information about the surroundings, e.g. color information that is especially useful to classify and categorize different objects. The major disadvantage of monocular cameras is on the one side the lack of depth information provided by the sensor and on the other hand, the high density of the data which requires heavy postprocessing to reason about the major quantity of interest: object location and motion make in the scene. This is achieved by object detection and tracking algorithms.

During the last years, typical object detection algorithms for monocular camera system like in static, classical computer vision applications techniques like Haar-Cascades have been completely replaced by Deep Learning based detectors such as Yolo-based Detector, RCNN-based detectors or recent Transformer-based Detection algorithms. Even 3D-Detectors such as FCOS3D are possible that overcome the problem that cameras only provide projective 2D information and hence overcome the lacking depth problem that monocular cameras lack depth information.

9.1.2 Road Side Stereo Camera (RSSC)

A stereo camera is equipped with two lenses and image sensors positioned side by side, mimicking the human visual system. The purpose of having two lenses or image sensors is to capture two slightly different perspectives of a scene, which allows for direct depth perception and the creation of three-dimensional (3D) images. Road side stereo cameras often use vertically displaced lenses and image sensors. This is advantageous for the following reasons:

- The camera needs gray value gradients, or edges, in the image to measure the stereo parallax, also called disparity. Road vehicles have many horizontal edges that support the measurement process.
- Road side cameras need a wide stereo basis. The longer the range of a stereo camera, the wider is the stereo basis needed. The wider stereo basis is easier to implement with a pole setup.

To create a 3D effect, the stereo camera uses called stereo vision or stereoscopy. By matching corresponding points in the different images, the camera can calculate the depth information of the scene. This depth information allows for the creation of a depth map or a point cloud, which represents the 3D structure of the captured scene.

The depth uncertainty of a stereo camera increases with object distance[12] and is highly dependent on parameters such as the visual appearance of the target.

9.1.3 Road Side Lidar (RSL)

Lidar sensors are active sensors that emit electromagnetic waves in the near infrared spectrum and typically use the time of flight principle to measure distances of the spatial surroundings. LiDAR sensors typically provide pointclouds as an intermediate data representation of all measured objects in the traffic scene and thus provides precise depth information of the surfaces of surrounding objects. To make use of lidar sensors processing algorithms are needed.

An advantage of lidar sensors is the precision of the measured pointcloud with a dense level of 3D information for high resolution scanners. A major disadvantage is that the denseness of the pointcloud depends on the distance of the objects to the sensor and the lack of visual information compared to camera sensors. This problem currently gets compensated by current lidar scanners that provide additional information like the near-infrared channel of the receiver to the application.

9.1.4 Road Side Radar (RSR)

This section is not filled in v1.0 of this document and will be detailed later, if required.

9.1.5 Road Side Thermal/Infrared Camera (RSIC)

Thermal imaging sensors measure a traffic scene in the infrared spectrum, providing reliable pixelwise intensity information that can be used for the recognition of vehicles, persons, animals and other traffic participants. Although computer vision algorithms from the RGB domain can be adapted to the infrared domain, thermal images impose a higher variety of visual properties depending on the traffic scene. E.g. insolation may create different thermal artefacts on streets and traffic objects, introducing challenges for detection algorithms in thermal images. Although being challenging in heavy sun, infrared cameras are able to detect persons during nighttime and have been widely used in military and police surveillance applications.

[12] Schreve, K. (2014, September). How accurate can a stereovision measurement be?. In *15th International Workshop on Research and Education in Mechatronics (REM)* (pp. 1-7). IEEE.

9.1.6 Road Side Communication

The Road Side Unit (RSU) is a critical hardware component that serves as a facilitator for vehicle-to-everything (V2X) communication. Its primary function is to establish information exchange between vehicles and the infrastructure, enabling advanced features such as collision avoidance, traffic management, and cooperative driving. In the context of the project, the RSU's interface plays a crucial role in connecting it with the EPU. This interface allows for efficient data transfer between the RSU and the EPU, also forwarded to the Backend, facilitating the forwarding of V2X data for remote operation.

While all V2X data holds relevance for remote operation, the encoded information regarding detected objects within the Cooperative Perception Message (CPM) assumes particular significance as it provides remote operators with situational awareness, allowing them to make informed decisions. Therefore, the RSU also collects V2X data of passing non-MAV vehicles.

9.1.7 Road Side Wifi Network.

The deployment of Wi-Fi access points is essential for enabling remote operation and connectivity for all non V2X usecases, esp. in the cases which require a larger bandwidth. These access points provide wireless network connectivity, allowing for low latency, high bandwidth data connection to the EPU and backend. This can be used for high bandwidth video streams and also additional data about the vehicle state that is normally not transmitted via V2X. These capabilities can also be handled by mobile network, with lower reliability and more delay to the backend, but with a much wider area coverage. For this reason, a mobile connection from the test vehicle to the remote operation station should also be established.

10 Annex 1 Glossary, Taxometry, Abbreviations

State of AD Definitions

In the automated driving community, a wide range of technological approaches are applied and so wordings, definitions, glossary, abbreviations are complex. In the European Project series with support actions for CCAM (Arcade, Fame etc.) consequently presented at the website www.connectedautomateddriving.eu tries to to standardize and reach harmonized use of terminology through classification such as a taxonomy[13]. However, development for AD is ongoing and so the Fame definitions are a good baseline but a living document. Fame Glossary is not consequent and abbreviations are not clear. Fame Taxometry is a very early approach of classification of the AD topic and defines only 10 topics as "classes" (AI, Data, Evaluation, Infrastructure, Operation, Regulation, Safety, Technology, User and Vehicle).

Furthermore, the terms and topics are strongly biased by the todays CCAM view of AD and does not fit to new developments or variants. E.g. the Fame Glossary definition for "Automated vehicle" = "Automated vehicle refers to a vehicle fitted with an automated driving system (ADS) that uses both hardware and software to perform dynamic driving tasks associated with moving the vehicle within a defined operational design domain (ODD).". This is a limited definition. It excludes vehicles which are not equipped with an ADS but interact with a ADS where the ADS is fitted in an edge computing unit in the infrastructure. In this case the definition has to be

[13] Glossary and Taxonomies - Connected Automated Driving ref. 2025-01-07

updated like: “Automated vehicle” = “Automated vehicle refers to a vehicle which is fitted with an automated driving system (ADS) or controlled by an ADS allocated in the infrastructure” (The rest is a definition of an ADS and another glossary topic).

MAD Definitions

This document tries to follow the Fame Glossary and Taxometry as far as possible. New or modified definitions are summarized in the following table:

Work in progress