European Commission

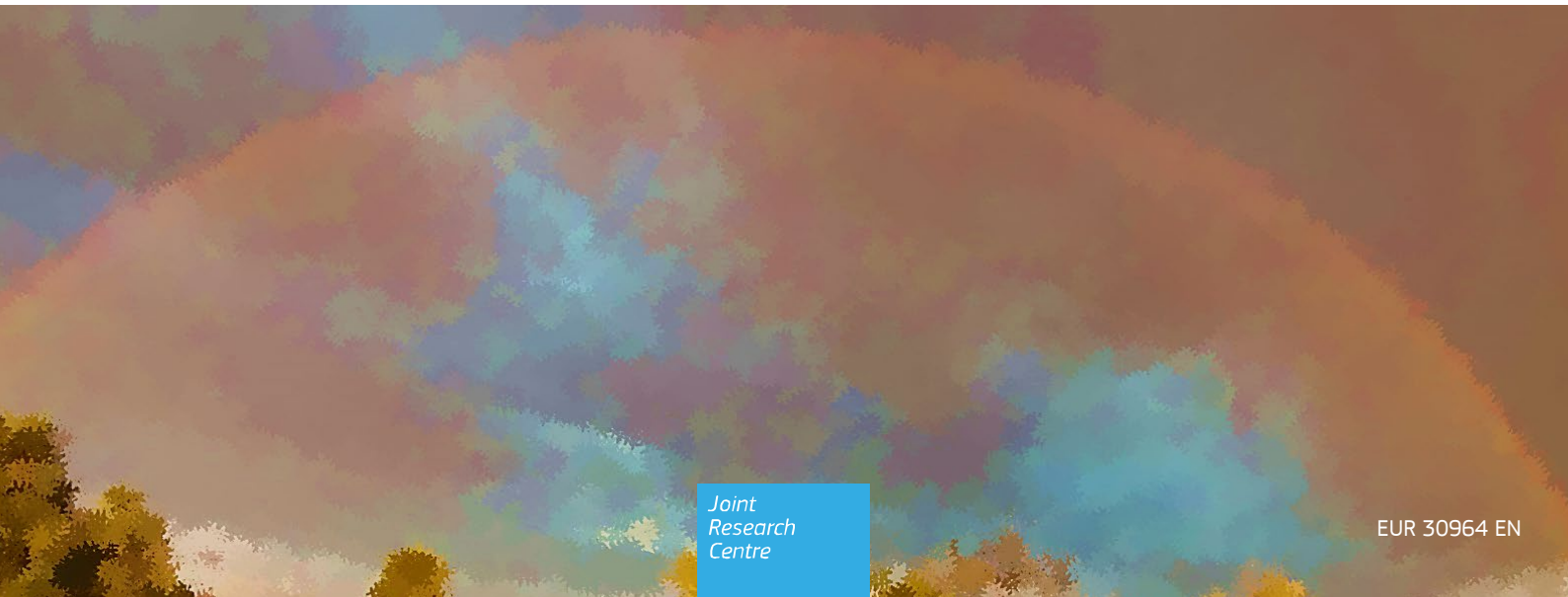
JRC SCIENCE FOR POLICY REPORT

Collection of fleet-wide fuel and energy consumption data from road vehicles

Investigation of a possible vehicle-to-cloud communication system for anonymous data collection

Kourtesis D., Fontaras G.

2022



Joint Research Centre

EUR 30964 EN

This publication is a Science for Policy report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the European Union concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Name: Georgios Fontaras
Address: European Commission, Joint Research Centre (JRC) Via E. Fermi 2479, TP-230
Email: georgios.fontaras@ec.europa.eu
Tel.: +39 0332 786425

EU Science Hub
https://ec.europa.eu/jrc

JRC127600

EUR 30964 EN

PDF ISBN 978-92-76-46614-7 ISSN 1831-9424 doi:10.2760/46158

Luxembourg: Publications Office of the European Union, 2022

How to cite this report: Kourtesis, D. and Fontaras, G., *Collection of fleet-wide fuel and energy consumption data from road vehicles*, EUR 30964 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-46614-7, doi:10.2760/46158, JRC127600.

Contents

Abstract

The JRC has been researching the technical challenges and solutions pertinent to collecting fleet-wide On-Board Fuel Consumption Measurement (OBFCM) data through direct data transfer from vehicles to the European Commission. In the scope of this research work, a technical solution framework was developed that conceptualizes a direct Over-The-Air (OTA) data transfer approach that is secure, privacy-preserving, tamper-resistant, scalable and future-proof. The report investigates the technical requirements regarding off-board, vehicle-to-cloud communication. The scope of this report is to introduce a solution architecture for the direct transfer of OBFCM data from vehicles to the EC and to demonstrate its feasibility by presenting a complete Proof-of-Concept implementation of this architecture built with open standards, modern web technologies and widely adopted cloud infrastructure services.

Acknowledgements

The authors would like to acknowledge the contribution to this study by European Commission colleagues Carlos Serra and Nikolaus Steininger from the Directorate-General for Climate Action (DG CLIMA), who provided invaluable input on challenges and requirements surrounding the collection of OBFCM data and for reviewing this report.

The authors are grateful to Athanasios Dimaratos from Aristotle University's Laboratory of Applied Thermodynamics for his support.

The authors would like to acknowledge the constructive remarks of the anonymous JRC reviewer who commented on the draft version of the report. The authors would like to thank Mrs Nadia Mesi for her support in the editing of this report.

Lastly, the authors would like to express their gratitude to independent experts Anastasios Valsamidis and Stefanos Doumpoulakis for their extensive contribution to the design and implementation of the proof-of-concept implementation described in this report.

Authors

Dimitrios Kourtesis (Ideas Forward)

Georgios Fontaras (European Commission, Joint Research Centre)

Executive summary

Policy context

The European Commission (EC) is investigating approaches and tools to monitor the real-world fuel consumption and CO_2 emissions of road vehicles in Europe. The aim is to ensure that the emissions reduction goals introduced by the European regulations are also reflected in vehicles' real-world operation. Regulation (EU) 2018/1832 introduced On-board Fuel and/or Energy Consumption Monitoring Devices (OBFCM devices) as a mandatory instrument in all new light passenger and commercial vehicles from January 2021 onwards. Subsequent CO_2 standards regulations gave the Commission the responsibility to collect and analyse the data recorded by OBFCM devices.

According to Article 12 of Regulation (EU) 2019/631, the Commission shall collect, from 2021, data on the real-world fuel or energy consumption of passenger cars and light commercial vehicles as measured and logged by OBFCM devices. The regulation outlines three different routes through which the Commission shall regularly be collecting the OBFCM measurement data: "*from manufacturers, national authorities or through direct data transfer from vehicles*".

Initially, the data will be collected through manufacturers and national authorities according to Commission Implementing Regulation (EU) 2021/392. The regulation introduced a legal requirement for manufacturers and the Member States to collect data from OBFCM devices every calendar year and report it to the Commission using the existing data exchange platforms provided by the European Environmental Agency (EEA). Manufacturers will be collecting the data either through direct data transfer from vehicles to their back-end systems or through their authorised dealers or authorised repairers when vehicles undergo routine maintenance or repairs. Member states will be collecting the data through the bodies or establishments responsible for roadworthiness testing when the vehicles undergo a Periodic Technical Inspection (PTI), an event that for light-duty vehicles usually takes place two to four years after it enters into circulation.

The investigation described in this report focuses on the technical requirements surrounding the third route of data collection envisaged: direct data transfer from vehicles to the European Commission, without any intermediating entities linked to the vehicles' lifecycle.

Scope of this study

The JRC has been researching the technical challenges and solutions pertinent to collecting fleet-wide OBFCM data through direct data transfer from vehicles to the Commission. In the scope of this research work, a technical solution framework was developed which conceptualizes a direct Over-The-Air (OTA) data transfer approach that is secure, privacy-preserving, tamper-resistant, scalable and future-proof. The study encompasses technical requirements regarding both on-board and off-board communication; however, this report focuses exclusively on off-board, vehicle-to-cloud communication.

The purpose of this report is to introduce a solution architecture for the direct transfer of OBFCM data from vehicles to the EC and to demonstrate its feasibility by presenting a complete Proof-of-Concept implementation of this architecture built with open standards, modern web technologies and widely adopted cloud infrastructure services.

Main outcomes

The report introduces an off-board communication protocol for OTA transfer of OBFCM data defined based on the MQTT protocol — a standard with growing adoption by the automotive industry (ISO/IEC 20922). The communication protocol proposed enables the integration of the three main technology components of the OTA data collection network: a) the OTA Devices, b) the OTA Registration Service, and c) the OBFCM Data Service. We discuss functionality, the context of use, frequency/volume of use, and operational supervision for each component.

In our definition of MQTT-based communication, we cover the MQTT topic schema and respective topic access rights, MQTT message types, and structure. Following, we present a proof-of-concept implementation of this solution architecture, developed to demonstrate its feasibility.

In presenting the PoC implementation, the report outlines the five different applications that make up the PoC components, i.e., OTA Registration Service, OBFCM Data Service, Fleet simulator, Certificate Authority simulator, and Workflow demonstrator. We discuss functionality, technology stack and implementation decisions. We also describe in detail the three most important processes carried out in the OTA data collection network and

implemented in the PoC, walking through these processes step by step with the help of sequence diagrams. The last parts of the report elaborate on how identity management and access management were realized on AWS services.

Key conclusions

The PoC implementation demonstrates that the solution architecture put forward is plausible. The study team developed a fully operational system that validates the solution approach from a functional and scalability viewpoint. The principal solution characteristics are:

- **Data reliability and privacy by design**, made possible through public-key cryptography (digital certificates, encryption, cryptographic signatures, PKI system) and tamper protection measures in OTA Devices (Trusted Platform Module).
- **Anonymous data collection**, achieved through pseudonymization of the data source (associating the collected OBFCM data with an OTA Device ID instead of the VIN) and data segregation (keeping vehicle data physically separate from vehicle identifiers).
- **Scalability to over 100 million vehicles**, made possible by extending the architecture with highly-scalable cloud infrastructure services (asynchronous message queueing, serverless Lambda functions, horizontal application scaling for concurrent processing).
- **Low communication footprint**, estimating the footprint of the OTA data communication to be less than 1MB per vehicle per year.
- **Manageable operational costs** considering the total cloud infrastructure costs for OTA data collection from 100 million vehicles.

Compared to data collection through manufacturers or Member States as introduced by 2021/392/EU, the collection of fleet-wide data through direct data transfer from vehicles has some distinct advantages:

- **Reduced lead-time for CO_2 monitoring**: Direct data transfer from vehicles to the Commission enables early detection of gaps between the reported and actual CO_2 emissions of a vehicle segment and early detection of data reporting problems This is because the direct OTA data collection mechanism is automatically initialized as soon as the vehicle leaves the production floor and is readily available to start the transfer of data. Such a possibility enables detecting problems within weeks from when a new vehicle enters into service – rather than months or years.
- **Higher data resolution**: The ability of the Commission to receive OBFCM data snapshots as frequently as necessary, and an eventual possibility of modifying the frequency of data acquisition dynamically, enables statistical analysis at a high level of resolution now and in the future. This will allow the Commission analysts and other involved stakeholders to observe fleet-scale patterns and understand the contributing factors to variability in CO_2 emissions (e.g. seasonality) in ways that had not been possible before.
- **Reduced data risk:** OTA data transfer directly to the EC reduces the risk of error, tampering, or unintended personal data exposure by minimizing the number of intermediate data processing systems/entities and eliminating human operator involvement.
- **Increased transparency**: Eliminating intermediaries and implementing a common OBFCM data collection mechanism for all manufacturers increases transparency and traceability. Future problems with data collection, such as data availability or data integrity issues, will be much easier to detect and address. Transparency makes root cause analysis much simpler compared to auditing the proprietary systems of the heterogeneous organizations involved in the data chain (manufacturers, dealers, repairers, PTI agencies, national authorities).
- **Enhanced data privacy**: Direct data transfer from vehicles to the Commission assumes a shorter data chain. It allows to minimize the number of intermediate data controllers and to universally and uniformly enforce personal data protection measures. It also provides one additional measure of data protection by allowing anonymization of the data collection process.

1 Introduction

Road transport contributes approximately 23% of the European Union's (EU) total carbon dioxide emissions (CO_2), the main greenhouse gas. The European Commission (EC) investigates more effective tools to track the actual CO_2 emissions produced by vehicles in Europe and monitor that the reductions foreseen by the binding CO_2 emission targets set by EU legislation are also reflected in real-world vehicle operation. The need for real-world vehicle emissions monitoring has led to new regulatory developments such as Regulation (EU) 2018/1832 [1], which introduced On-board Fuel and/or Energy Consumption Monitoring Devices (OBFCM devices) as a mandatory feature of every new Light-Duty Vehicle (LDV) from January 2021 on.

1.1 On-board fuel and electric energy consumption monitoring in new vehicles

According to Regulation (EU) 2018/1832, the term OBFCM device means "any design element, either software and/or hardware, which senses and uses the vehicle, engine, fuel and/or electric energy parameters". An OBFCM device shall determine at least the following parameters and store the lifetime values on board the vehicle:

- total fuel consumed (lifetime) [litres]
- total distance travelled (lifetime) [kilometres]
- engine fuel rate [grams/second]
- engine fuel rate [litres/hour]
- vehicle fuel rate [grams/second]
- vehicle speed [kilometres/hour]

Specifically for hybrid electric vehicles that can be charged from an external source (Off-Vehicle Charging Hybrid Electric Vehicle), OBFCM devices will also monitor and store the following values:

- total fuel consumed in charge depleting operation (lifetime) [litres]
- total fuel consumed in driver-selectable charge increasing operation (lifetime) [litres]
- total distance travelled in charge depleting operation with engine off (lifetime) [kilometres]
- total distance travelled in charge depleting operation with engine running (lifetime) [kilometres]
- total distance travelled in driver-selectable charge increasing operation (lifetime) [kilometres]
- total grid energy into the battery (lifetime) [kWh]

Post-2020 CO_2 emission targets for light and heavy-duty vehicles have been established by Regulations (EU) 2019/631 [2] and (EU) 2019/1242 [3], respectively. The same regulations, mandate the European Commission (EC) to perform the monitoring describing the management of the data:

(a) The Commission shall process the collected data and create anonymised and aggregated datasets, including per manufacturer;

(b) The aggregated data shall also be made public to show how the real-world representativeness evolves;

(c) The Commission will have to ensure that the VIN data are not kept longer than is necessary according to the EU data privacy provisions.

The regulation introduces the requirement for anonymised aggregated datasets collected per OEM basis that are later published to demonstrate the representativeness of real-world CO_2 emissions. The regulation also highlights the EU data privacy provisions, a significant concern for many vehicle owners and users.

1.2 Enabling fleet-wide analysis of CO_2 emissions through Over-the-Air data collection

There are multiple routes through which the EC could collect OBFCM data to perform fleet-wide analysis and to monitor important aggregate metrics relating to emissions. According to Article 12 of Regulation (EU) 2019/631, the Commission shall collect, from 2021, data on the real-world fuel or energy consumption of passenger cars and light commercial vehicles as measured by OBFCM devices. The regulation outlines **three different routes**

through which the Commission shall regularly be collecting the OBFCM measurement data: ***"from manufacturers, national authorities or through direct data transfer from vehicles"***.

How data will be collected via the first two routes, i.e., through manufacturers and national authorities, has been set in Commission Implementing Regulation (EU) 2021/392. The regulation introduced a legal requirement for manufacturers and the Member States to collect data from OBFCM devices every calendar year and report it to the Commission using the existing data exchange platforms provided by the European Environmental Agency (EEA). Manufacturers will be collecting the data either through direct data transfer from vehicles to their back-end systems or through their authorised dealers or authorised repairers when vehicles undergo routine maintenance or repairs. Member states will be collecting the data through the bodies or establishments responsible for roadworthiness testing when the vehicles undergo a Periodic Technical Inspection (PTI), an event that for light-duty vehicles usually takes place two to four years after it enters into circulation. In both cases, data collection will not take place if the vehicle owner expressly refuses to make data available.

The present study focuses on the third route of data collection foreseen by Regulation (EU) 2019/631: *direct* data transfer from vehicles to the European Commission, i.e. data transfer from vehicle systems directly to Commission systems, without any intermediating entities.[1]

Compared to data collection through manufacturers or the Member States as mandated in Regulation (EU) 2021/392, the collection of fleet-wide data through direct data transfer from vehicles has some distinct advantages:

- **Reduced lead-time from vehicle registration to CO_2 monitoring**: Collecting OBFCM data directly from the vehicle will allow the Commission to obtain reliable statistical figures within weeks from the date a new vehicle type enters the market. The lead-time from vehicle registration to collecting the first OBFCM data is reduced considerably compared to collecting data via manufacturers or via the Member States and the PTI process. This, in turn, allows data reporting problems to be detected early enough to mitigate their impact. It also allows CO_2 limits per manufacturer to be monitored more accurately and fairly.
- **Higher data resolution**: Periodic data transfer directly from the vehicle to the Commission allows data acquisition to take place as frequently as needed to understand the effect of seasonal phenomena on fuel and/or energy consumption, including seasonal differences at the country level. This is a departure from existing plans to collect OBFCM data every time a vehicle visits a roadworthiness test facility or a manufacturer's authorized repairer.
- **Reduced data risk**: OTA data transfer directly to the EC means there is a short data chain without intermediaries storing, processing or forwarding the OBFCM data. It eliminates human operators. This reduces the risk of errors, the risk of data tampering (by manufacturers or others), and the risk of unintended personal data exposure (as per Commission Implementing Regulation (EU) 2021/392, Vehicle Identity Numbers are considered personal data from the moment the vehicle is registered).
- **High transparency**: Eliminating intermediaries and implementing a common OBFCM data collection mechanism for all manufacturers increases transparency and traceability. Future problems with data collection, such as data availability or data integrity issues, will be much easier to detect and address. Enhanced transparency makes root cause analysis much simpler compared to auditing the proprietary systems of the heterogeneous organizations involved in the data chain (manufacturers, dealers, repairers, PTI agencies, national authorities).
- **Enhanced data privacy**: Direct transfer of encrypted data from vehicles to the Commission makes it possible to not only minimize the number of intermediate data controllers but, most importantly, to universally and uniformly enforce personal data protection measures. It provides one additional measure to strengthen data privacy through pseudonymization of Vehicle Identity Numbers.

In direct OTA data collection, OBFCM data is transferred from the vehicle to a special-purpose cloud infrastructure under the European Commission's responsibility. Connectivity between the vehicle and the cloud is accomplished via the mobile network. Data transfer happens automatically while the vehicle is under normal operation, without requiring any action or effort by the driver or other person.

[1] One could argue that a scheme where the vehicle manufacturer's systems collect OBFCM data from vehicles Over-The Air, and subsequently relay this data to the Commission, would be a form of quasi-direct transfer. Such a transfer scheme, however, would depend on the vehicle manufacturer operating as intermediary and have not been analysed in the present report.

The frequency at which the data transfer takes place could be adjustable, ranging anywhere between once per month and once per year. A single data transfer could include multiple periodic snapshots of the OBFCM parameter values, such as once per month to once per quarter. This flexibility is another benefit of the direct data transfer route: the frequency at which OBFCM value snapshots are taken and transferred can be dynamically modified, offering optimal resolution in vehicle types of high interest, e.g. more frequent sampling in more recent vehicles compared to older models. Such quick adaptations can save resources and help better understand the emission and fuel consumption patterns in different segments of the fleet. The objective, of course, is not to monitor segments of one – i.e., individual vehicles, but large sets of vehicles from which aggregate fuel and energy consumption metrics can be derived.

1.3 The role of the 'OTA Device' in the end-to-end solution framework

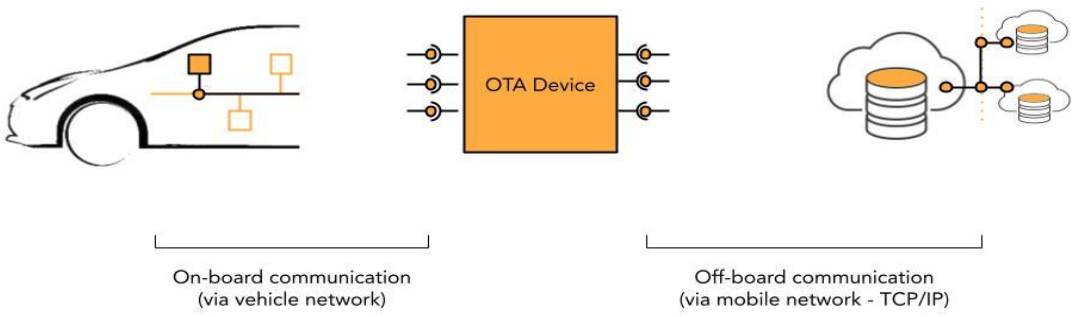
The JRC has been researching to analyse the requirements and investigate solutions for the challenge of real-world vehicle emissions monitoring based on OTA transfer of data directly from the vehicle. In this context, a technical solution framework was developed, which was first presented at the 6th OBM Task Force Meeting on 15 December 2020 [4]. The objective of the proposed framework was to conceptualize an end-to-end solution approach for OTA transfer of OBFCM data that is secure, privacy-preserving, scalable, tamper-resistant and future-proof.

1.3.1 The OTA Device

The solution framework introduces the concept of an OTA Device, i.e., of an on-board software or hardware unit which is simultaneously integrated with both on-board and off-board systems:

- **On-board integration**: The OTA Device communicates via the vehicle network with the on-board OBFCM device and other on-board ECUs if needed
- **Off-board integration**: The OTA Device communicates via the mobile network to cloud services which are under the responsibility of the European Commission

Figure 1. Conceptual illustration - OTA Device on-board and off-board integration



Source JRC 2022

1.3.2 OTA Device states

We envisage six different high-level states in which the OTA Device may be found during its regular operation:

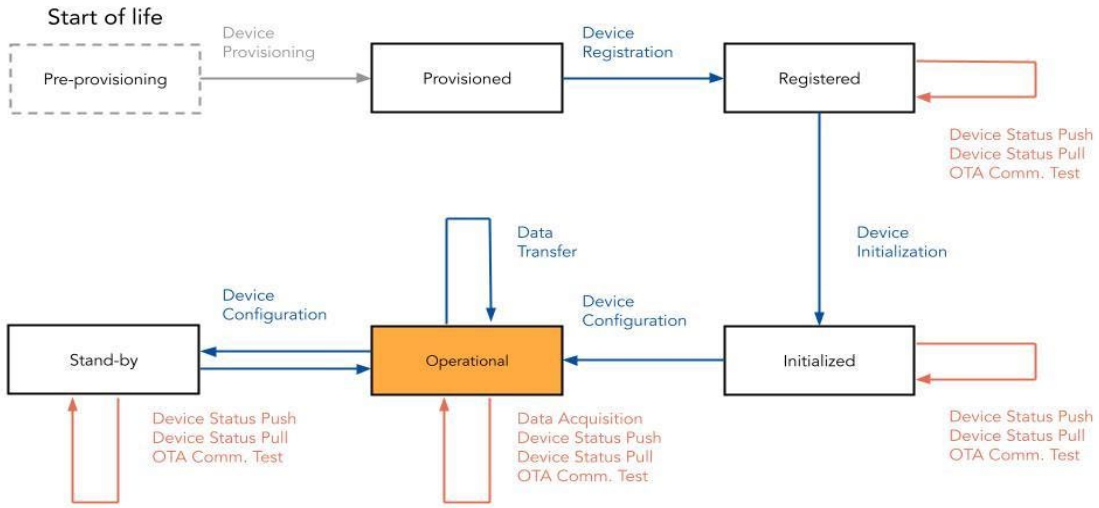
1. **Pre-provisioning state**: Off-board communication is not yet possible in this state because the OTA Device does not yet have access to some of the necessary security-related assets (such as digital certificates, security keys, access credentials or usage tokens) or because necessary configuration settings are still missing. Being in this state means that the device installation is not complete.
2. **Provisioned state**: In this state, all necessary security-related assets and factory configuration parameters have been provisioned to the OTA Device. The OTA Device has been successfully installed on a vehicle and can communicate with off-board systems via the mobile network. However, the device has not yet registered itself with the off-board system responsible for admitting the device into the OTA data collection network.

3. **Registered state**: In this state, the device has successfully registered with the relevant system responsible for OTA Device registration. It has been admitted to the OTA data collection network. However, the device's configuration parameters have not yet been initialized. The OTA Device will communicate with the OBFCM data collection service as soon as it can connect to the internet.
4. **Initialised state**: In this state, the OTA Device has successfully published an initialisation request to the remote OBFCM data collection service and has successfully obtained instructions to initialise the device's configuration parameters. Being in this state means that the vehicle on which the OTA Device has been installed has been registered as a new vehicle in the EU, but additional legal requirements may not be fulfilled (e.g., vehicle owner consent may not have been granted if applicable). The OTA Device is waiting until it becomes allowed to start sending data.
5. **Operational state**: In this state, the OTA Device periodically retrieves data from the OBFCM device (acquiring snapshots of OBFCM values) and, later in time, transfers this data to the remote data collection service. Being in this state means that the vehicle on which the OTA Device has been installed has been registered as a new vehicle in the EU, and any additional legal requirements are fulfilled (e.g., vehicle owner consent was granted, if applicable).
6. **Stand-by state**: In this state, communication is placed on hold. The device has obtained and executed instructions to cease reporting OBFCM data to the remote data collection service until further notice.

1.3.3 OTA Device functions

The model diagram in Figure 2 below illustrates the six states mentioned above of the OTA Device as well as the possible transitions from one device state to another. Transitions are labelled with the names of nine OTA Device functions. When executed, these nine functions move the OTA Device from one state to the next. Functions that involve off-board communication are shown in blue font whereas functions involving on-board communication are shown in red.

Figure 2: Model of the OTA Device functionality - state transitions labelled by functions



Source JRC 2022

Table 1 follows briefly describes each of the nine OTA Device functions. The first column indicates the type of communication interface which is exercised by each function.

Annex 1 provides a table showing the source and destination of information flows taking place when each device function is exercised, in both on-board and off-board communication.

Table 1. OTA Device functions

Interface	Function	Function description
Off-board communication	Device Registration	The OTA Device communicates with a remote device registration/management service to present its digital certificate and to register on the OTA data collection network.
	Device Initialization	Once successfully registered, the OTA Device communicates with the offboard data collection service to receive its initial configuration settings.
	Device Configuration	Once successfully initialized the OTA Device communicates with the data collection service to receive updated configuration settings, including instructions to potential transition to a new state.
	Data Transfer	The OTA Device communicates with the offboard data collection service to transfer the pre-stored OBFCM parameter values.
Both On-board and Off-board (depends on implementation)	Device Provisioning	The OTA Device initiates any necessary operations to bring itself into a state where all provisioning artefacts such as digital certificates and factory configuration settings are in place. This involves communicating with offboard and/or onboard systems. How this is done depends on specific implementation decisions.
On-board communication	Data Acquisition	The OTA Device communicates with the onboard OBFCM device to retrieve OBFCM parameter values and store them locally.
	Device Status Push	The OTA Device communicates with onboard ECUs to notify them about an internal state change or an important event (e.g., error).
	Device Status Pull	Onboard ECUs (e.g., HMI) or OBD connected devices (a scan tool) communicate with the OTA Device to query its internal state and/or to retrieve the device event log.
	OTA Communication Test	Onboard ECUs (e.g., HMI) or OBD connected devices (a scan tool) communicate with the OTA Device to trigger a manual communication cycle for diagnostics purposes such as during Periodic Technical Inspection.

Source: JRC, 2022

1.4 Data collection risks affecting the reliability of OBFCM data

Any technical solution approach for OTA transfer of OBFCM data must be evaluated concerning the measures it puts in place, by design, to guard against data reliability and privacy risks. Safeguards cannot be an afterthought. The critical risks, in the OBFCM data collection use case, are:

- **Data availability risk**: OBFCM data from specific vehicles is not received regularly or not received at all.
- **Data integrity risk**: OBFCM values are altered/tampered with before or during data transfer.
- **Data authenticity risk**: The data are not truly sourced from the supposed vehicle.
- **Data confidentiality risk**: Personal or non-personal data are exposed to unauthorized entities.

We argue that implementing the OTA Device as a standalone, single-purpose unit that shares as little resources as possible with other systems and doesn't involve intermediaries will increase transparency, increase security and tamper-resistance and provide stronger safeguards for data privacy.

1.5 Direct OTA data transfer approaches

When introducing our proposed solution framework for OTA transfer of OBFCM data [4] we considered alternative approaches for implementing the functionality of the OTA Device and integrating it into the vehicle environment. Transport sector associations such as ACEA, AIRC, CECRA, FIA and FIGIEFA have further proposed their solution approaches for OBFCM data collection, looking at the matter from their perspective and according to their priorities.

The various approaches can be compared in terms of two critical dimensions:

- **Direct vs indirect communication**: Is the data transfer from the vehicle to the EC happening through a direct communication link between the two or indirectly through intermediate systems?
- **Shared vs dedicated on-board resources**: Does the operation of the OTA Device relies on shared or dedicated resources of the vehicle? Key resources include the on-board software execution environment, security/cryptography processes and assets (e.g., digital certificates and secure storage for secret encryption keys), mobile network connectivity resources (for access to the internet) and vehicle network interfaces (for access to the OBFCM data).

Those two dimensions are critical because they determine the transparency afforded by each solution and its assurances against data reliability and privacy risks.

Table 2 below compares the solution approaches for direct transfer of OBFCM data discussed to date, in conversations between the authors and various transport sector stakeholders and also in the context of the European Commission's OBM Task Force.

Table 2. Integrating the OTA device functionality into the vehicle - Classification of approaches

On-board integration approach	Description	Direct or indirect data transfer	Shared or dedicated on-board resources
Self-contained hardware unit relying on private resources.	The OTA Device functionality is implemented in the form of a standalone, single-purpose embedded control unit containing all necessary resources for its operation, including private mobile network connectivity for access to the internet. The hardware unit is produced by Tier 1 Suppliers who are also responsible for Type Approval of the unit and software updates. In [4] we referred to this concept as "White-box OTA Device".	Direct transfer (vehicle to EC)	Relying exclusively on dedicated on-board resources (computing platform, cryptoprocessor, connectivity (4G modem + eSIM + mobile plan), vehicle network interfaces).
Special-purpose hardware unit relying on some shared resources.	The OTA Device functionality is implemented in the form of a standalone, single-purpose embedded control unit that does not have its own built-in mobile network connectivity. The unit relies on the vehicle's existing mobile connection to connect to the cloud. The OTA Device is produced by Tier 1 Suppliers who are also responsible for Type Approval of the unit and software updates. In [4] we referred to this concept as "Grey-box OTA Device".	Direct transfer (vehicle to EC)	Relying mostly on dedicated onboard resources. On-board resources for connectivity (4G modem + eSIM + mobile plan) are shared with other vehicle components. Depending on implementation it may also rely on shared off-board resources such as the OEM's private computer network.
Special-purpose software unit deployed on a shared multi-purpose on-board software platform.	The OTA Device functionality is provided by a software service/application which is deployed on a shared multi-purpose "sandbox" execution environment. The app is developed by the vehicle manufacturer who is also responsible for software updates. The software unit which realizes the OTA Device functionality is distinct from other software subsystems and can be independently updated. It is unclear if the unit undergoes Type Approval as a separate automotive component. FIA, FIGIEFA, AIRC, CECRA have proposed a concept along these lines referred to as "Secure On-board Telematics Platform" [5].	Direct transfer (vehicle to EC)	Relying exclusively on shared on-board and off-board resources. Shared on-board: computing platform, cryptoprocessor, connectivity, vehicle interfaces are all shared). Shared off-board: OEM's private computer network.
Software logic integrated into multi-purpose on-board software.	The OTA Device functionality is implemented as a software module that is not distinct from other software deployed on the same ECU, in the sense that the specific functionality cannot be modified/updated independently of other software. Instead, it is part of a larger software subsystem. Updates come into effect as part of updating the larger software subsystem that the OTA logic is part of. The software is developed by (or under the responsibility of) the vehicle manufacturer, who is also responsible for any updates. It is unclear how Type Approval is scoped.	Direct transfer (vehicle to EC)	Relying exclusively on shared on-board and off-board resources.
No on-board integration	This approach does not involve on-board integration but is added here for comparison. The OBFCM data is collected via OEM's proprietary OTA mechanism and later in time made available in batches to the EC via an OEM back-end server. This is the "Extended Vehicle Server" concept proposed by ACEA [6]. The Extended Vehicle Server concept proposed by ACEA is not comparable to direct OTA. The OTA Device functionality is not implemented on-board the vehicle and there is no scope for Type Approval.	Indirect transfer (vehicle to OEM, then OEM to EC)	Relying exclusively on shared off-board resources.

Source: JRC, 2022

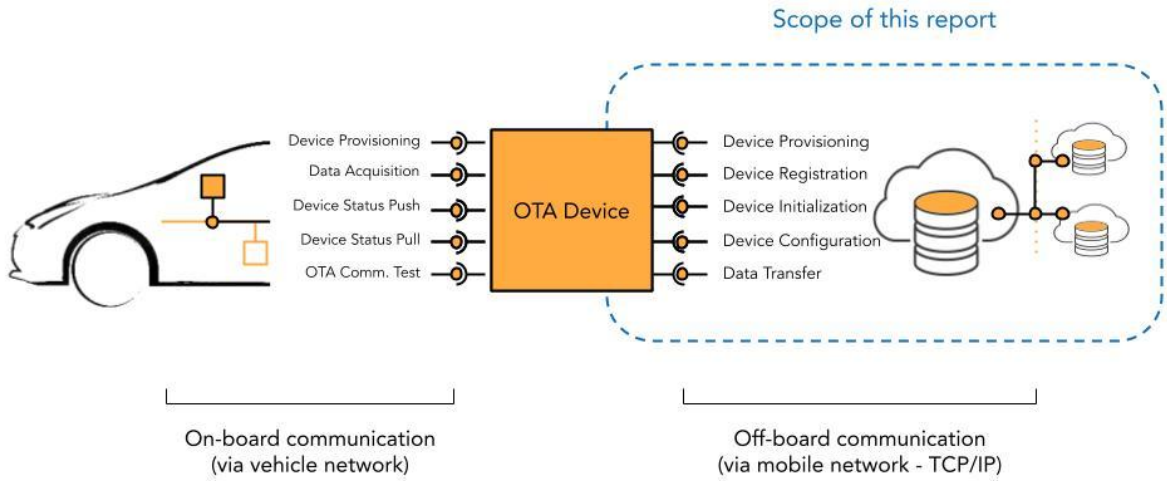
1.6 Off-board communication

The main focus of this report is off-board communication. The decision between a shared-resource vs a dedicated-resource model for implementing the OTA Device is very important, as it comes with significant trade-offs and implications. However, further analysis is not in the scope of this report.

In the following sections, we maintain an implementation-agnostic view of the OTA Device and its on-board integration model. We do not assume whether the OTA Device has been implemented in the form of a dedicated onboard controller unit or as a software module deployed on a multi-purpose controller unit (e.g., Engine Control Unit or Telematics Control Unit).

As illustrated in Figure 3 below, the focus of this report is on off-board integration, i.e., communication between the device and remote cloud services operated by the EC. The off-board communication protocol can be developed without making any assumptions about how the OTA Device communicates with the vehicle's on-board systems. The two problems are orthogonal.

Figure 3: Conceptual diagram illustrating on-board and off-board OTA Device interfaces



Source JRC 2022

Specifically, this report describes a communication protocol for vehicle-to-cloud communication built on the MQTT standard [7]. It showcases the implementation of this protocol using modern web technologies and scalable cloud infrastructure services, indicatively in this case provided by Amazon Web Services (AWS).

The next sections of this report present:

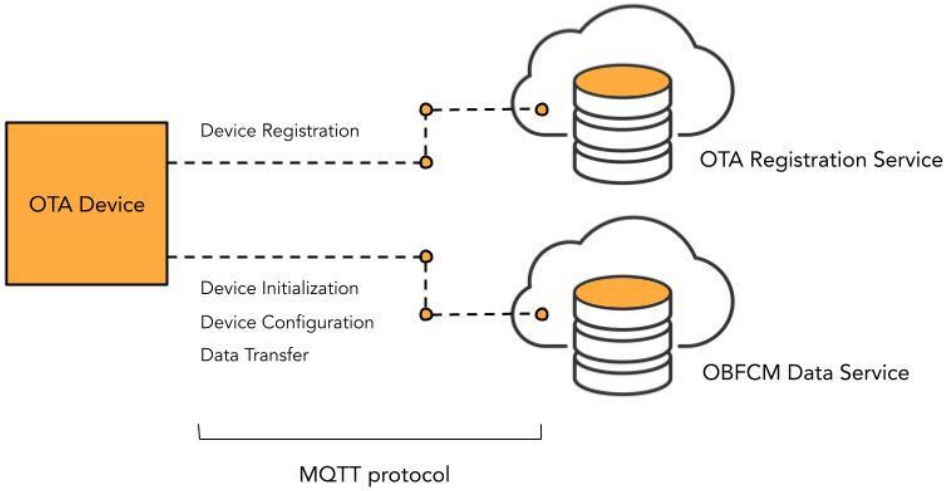
1. An overview of the off-board systems in the solution framework
2. A description of the off-board communication protocol
3. A description of the proof-of-concept implementation
4. A discussion on how to scale up OTA data collection
5. Conclusions and notes on further research

2 Off-board cloud services

As illustrated in Figure 4 below, the OTA Device engages in off-board communication with two remote cloud services:

- OTA Registration Service: Cloud service acting as a gateway that restricts enrolment to trusted OTA Devices only, enforces data collection policies and anonymizes the data collection process.
- OBFCM Data Service: Cloud service that receives OBFCM data from trusted OTA Devices previously registered with the OTA Registration Service. It also provides up-to-date configuration settings to the devices.

Figure 4: Off-board systems interfacing with the OTA Device



Source: JRC 2022

The focus of the technical analysis in this report is the communication between those two off-board systems and the OTA Device. The following subsections 2.1 and 2.2 introduce the OTA Registration Service and the OBFCM Data Service.

2.1 OTA Registration Service

2.1.1 What purpose does the system serve?

The OTA Registration Service is a system of record on the cloud where some basic information about every new vehicle entering circulation in the EU is securely entered to enable fuel and energy consumption monitoring across the entire EU fleet. This basic information includes the vehicle's unique identification number (VIN), technical vehicle specifications relevant to emissions monitoring (e.g., vehicle manufacturer, type approval number, engine type) and information about the OTA Device installed on the vehicle (most importantly, the OTA **Device's X.509 Digital Certificate**).

The service exists to allow new vehicles to be automatically enrolled into the **OTA data collection network**. Firstly, it acts as a gatekeeper that only admits OTA Devices presenting Digital Certificates from known and trusted sources. Secondly, it serves as a **policy enforcement gateway**, ensuring that data collection is performed only if the necessary preconditions are met. For instance, even if a vehicle's OTA Device presents trusted credentials and is accepted to join the OTA data collection network, the device will not be allowed to send OBFCM data until later in time, after the vehicle becomes registered in an EU country and, if applicable, after the vehicle owner has granted consent.

Once a new OTA Device successfully registers itself, the OTA Registration Service "introduces'' the device to the **OBFCM Data Service**. This allows the OTA Device to start communicating with the OBFCM Data Service directly. At the same time, the OTA Device is added to a temporary "**deny-list**", meaning that it is not allowed to send OBFCM data. Weeks or months later, when the vehicle is eventually sold in an EU country and data subject consent has been granted (if applicable), the OTA Device is moved from the deny-list to the "**allow-list**". The next time the OTA Device communicates with OBFCM Data Service, it will be instructed to start sending OBFCM Data according to a schedule.

The OTA Registration Service also serves the purpose of facilitating anonymous data collection. Anonymization is achieved by decoupling the identity of the actual vehicle (VIN) from the identity of the OTA Device (Digital Certificate ID) when collecting data. In this way, data are segregated. The OTA Registration Service knows **which VINs map to which OTA Device IDs,** but this information is not shared with the OBFCM Data Service. Conversely, the OBFCM Data Service knows the fuel/energy consumption and mileage data associated with each OTA Device ID but this information is not shared with the OTA Registration Service. This is a form of separation of duties based on the principle of least privilege (PoLP). It facilitates data segregation which elevates data privacy protection.

2.1.2 When is the system used?

The OTA Registration Service engages in action on the following events:

- When a new vehicle is produced: The OTA Device installed in a new vehicle that is manufactured in Europe will automatically communicate with the OTA Registration service as soon as it can connect to a mobile operator network, possibly even before the vehicle leaves the production floor. The OTA Device will present its Digital Certificate to the OTA Registration Service and will be enrolled in the OTA data collection network. However, the OTA Device will not instantly be allowed to send OBFCM data. Instead, it will be added to a waiting list (the "deny-list").
- When a new vehicle is sold and registered in a Member State: A back-end system operated by the Vehicle Registrar organization (vehicle manufacturer or Member State transport authority) will communicate with the OTA Registration Service to register the vehicle VIN and its technical specifications. Provided that all additional preconditions are met (for instance, data subject consent is granted, if applicable), the OTA Registration Service will move the OTA Device from the "deny-list" to the "allow-list" and will notify the OBFCM Data Service. This will initiate the periodic data collection cycle.

2.1.3 How often is it used?

Based on 2019 statistics by ACEA (pre-Brexit and pre-pandemic), approximately 1.3 million new passenger cars were produced in Europe per month [8], and 1.5 million new vehicles were registered per month [8].

This would translate to the following volume of communication events for the OTA Registration Service:

- Incoming MQTT messages due to new vehicles being produced: 1.4 million MQTT connections on average per month, coming from new vehicles' OTA Devices requesting registration with the OTA data collection network.
- Outgoing HTTP messages due to new vehicles being produced: 1.4 million HTTP connections on average per month to the OBFCM Data Service as notifications for new OTA Device registrations
- Outgoing MQTT messages due to new vehicles being produced: 1.4 million MQTT connections on average per month to send registration confirmations to OTA Devices.
- Incoming MQTT messages due to new vehicles being registered in a Member State: 1.5 million MQTT connections on average per month, coming from back-end systems of Vehicle Registrars (Member States or Vehicle Manufacturers, to be determined), assuming that all newly sold vehicles are "allow-listed" for data collection.
- Outgoing HTTP messages due to new vehicles being registered in a Member State: 1.5 million HTTP connections on average per month to the OBFCM Data Service as notifications for new vehicle registrations.

These communications total an average of 2.9 million incoming and 1.4 million outgoing MQTT communication events per month plus 2.9 million HTTP outgoing HTTP connections per month. Seasonal variability is expected, but also variability across days of the week and hours of the day.

2.1.4 Who is responsible for system operation?

Vehicle OTA registration could be seen under the broader context of vehicle digital identity attribution, and hence it could follow any such relevant developments in the future. It could be possible that the OTA Registration Service is maintained by the European Commission (EC) for all vehicles registered in the EU. Alternatively, each Member State could maintain a separate instance of this system for all vehicles registered nationally. In the latter case it will be necessary for the systems to be networked because the country in which a new vehicle is manufactured is not necessarily the country in which it is sold. Both approaches require high-availability guarantees and the ability to scale (proportionally to the anticipated volume of transactions per country).

2.2 OBFCM Data Service

2.2.1 What purpose does the system serve?

The OBFCM Data Service is a cloud-based system responsible for reliably receiving and storing the OBFCM data sent to the EC by tens of millions of vehicles across the EU-wide fleet. It provides a highly-scalable communication endpoint and large-scale storage suitable for time series data. The data remains encrypted both during transfer and when at rest.

In addition to receiving and storing the OBFCM data, this system is also responsible for sending instructions to individual OTA Devices about a) the frequency at which they should be acquiring OBFCM data readings from the vehicle's OBFCM device (e.g. once per month or once per quarter), b) the frequency at which they should be sending their acquired data to the OBFCM Data Service (e.g. once per quarter or once per year) and also c) the frequency at which they should be querying the OBFCM Data Service for updates to their configuration settings (e.g. once per month or once per quarter).

2.2.2 When is the system used?

The OBFCM Data Service engages in action on the following events:

- When a new vehicle is produced: After an OTA Device has successfully registered with the OTA Registration Service, the next step is to communicate with the OBFCM Data Service and check if there is any update that it should apply to its configuration settings.
- When a new vehicle is sold and registered in a Member State: Once the OTA Registration Service moves an OTA Device from the "deny-list" to the "allow-list", it notifies the OBFCM Data Service. Subsequently, the OBFCM Data Service prepares a configuration update for the specific OTA Device.
- When an OTA Device connects to request a configuration update: Every OTA Device periodically connects with the OBFCM Data Service to check if there is any update that it should apply to its configuration settings. The initial frequency at which this "check for updates" is happening is configured as a factory default.
- When an OTA Device connects to send OBFCM data: Every OTA Device periodically connects with the OBFCM Data Service to send the data it has already acquired from the vehicle's OBFCM device. The frequency at which this "data transfer" should happen is set by the OBFCM Data Service as soon as the OTA Device is moved from the "deny-list" to the "allow-list".

2.2.3 How often is it used?

The frequency at which the data transfer takes place could be dynamically controlled by the EC and could indicatively be once per quarter year[2], per vehicle. According to the 2021 edition of ACEA's 'Vehicles in use'

[2] For simplicity in this report we assume the same sampling frequency for all vehicles throughout their lifetime. Dynamic adjustment of sampling frequency could allow refined sampling that would improve the quality of the monitoring data. One could assume a more frequent data communication could be applied to newer vehicles, and a less frequent sampling to older vehicles. This would be a reasonable approach as the lifetime fuel consumption measurement becomes more robust when mileage and consumption accumulate, and the

report [9], the number of passenger cars on European Union roads reached 242.7 million while the number of light commercial vehicles reached 28.1 million. This makes a total of 270.8 million Light-Duty Vehicles.

Therefore, assuming a total EU wide fleet of 270 million LDVs which are all equipped with an OBFCM device and also with an OTA Device in an operational state, and assuming that the OTA Devices are configured to communicate with the OBFCM Data Service four times a year throughout their lifetime, the result is over 1 billion incoming data transfer events per year, or an average of 90 million MQTT messages incoming per month.

In addition:

- Incoming HTTP messages due to new vehicles being produced: 1.4 million HTTP connections on average per month, coming from the OTA Registration Service as notifications of new OTA Device registrations.
- Incoming MQTT messages due to new vehicles being produced: 1.4 million MQTT connections on average per month, coming from new vehicles' OTA Devices requesting initial configuration settings, after successful registration with the OTA Registration Service.
- Outgoing MQTT messages due to new vehicles being produced: 1.4 million MQTT connections on average per month, to initialize newly registered OTA Devices, by sending the latest configuration. This figure assumes that all OTA Devices will require an initial configuration setting that is different from their factory defaults.
- Incoming HTTP messages due to new vehicles being registered in a Member State: 1.5 million HTTP connections on average per month, coming from the OTA Registration Service as notifications of new vehicle registrations.
- Outgoing MQTT messages due to new vehicles being registered in a Member State: the 1.5 million HTTP notification messages per month from the OTA Registration Service will be followed by an equal number of MQTT messages for configuration updates which will instruct the respective OTA Devices of the newly-registered vehicles to transition to the "initialized" state.
- Incoming MQTT messages due to OTA Devices checking for new configuration updates: assuming config checks happening indicatively once per quarter year, there are 1 billion incoming config check events per year or an average of 90 million events per month.

This sums to an average of 181.4 million incoming and 2.9 million outgoing MQTT connections, plus another 2.9 million HTTP connections. The OBFCM Data Service must therefore have high-availability guarantees and be able to scale to accommodate peak demand.

Since data transfer operations will happen only if the vehicle is operational, variability across days of the week and hours of the day should be expected. The probability of a data transfer operation happening during working hours and days is more likely than on weekends and after hours when vehicle usage drops.

2.2.4 Who is responsible for system operation?

The operation and maintenance of the OBFCM Data Service are expected to be under the responsibility of the European Commission.

variability of the monitored values is less influenced by external factors. Also, older vehicle models will be probably of lower importance for monitoring purposes.

3 Off-board communication protocol

The definition of the off-board communication protocol comprises:

— **MQTT topic schema**: How the MQTT standard (ISO/IEC 20922:2016) is applied to enable the exchange of messages between OTA Device and off-board cloud systems, through the use of specific MQTT publish/subscribe topics.

— **Message types**: Which types of MQTT messages are exchanged between a) the OTA Device and the MQTT broker of the OTA Registration Service and b) the OTA Device and the MQTT broker of the OBFCM Data Service.

— **Message structure**: What is the message content structure, i.e., the data types included in the messages being exchanged.

The following sections of this chapter describe each of the above aspects.

3.1 MQTT topic schema & access rights

MQTT is a Publish/Subscribe messaging transport protocol. It is lightweight, open, simple, and designed to be easy to implement [10]. Heterogeneous systems can communicate by publishing MQTT messages to an MQTT broker, which acts as an intermediary to the communication. An MQTT topic is a hierarchically-structured character string denoting the subject of the communication. This character string is used to filter and route messages. Systems implementing the MQTT protocol use the topic of a message to determine which published message goes to which client (subscriber) [11].

How the MQTT topic schema is defined is an important architecture level decision because it affects the system's scalability, maintainability, and extensibility. As described in [12]: “MQTT topics must balance current device communications, cloud-side operations, and future device capabilities. Therefore, it can be challenging to design an ideal MQTT topic structure that creates enough of a schema to enforce least privilege communication but does not create a rigid structure that makes it challenging to support future device deployments.”

3.1.1 Topic schema

Device registration topics

We use two topics for the communication between the OTA Device and the MQTT broker of the Registration Service. One topic is for the device to publish messages which are received by the server and another for the server to publish messages received by the device:

- Topic `iot4emissions/registerDevice/<device-ID>` is where a device publishes registration requests
- Topic `iot4emissions/registerServer/<device-ID>` is where the OTA Registration Service publishes responses to the requests.

Device initialization & configuration topic

To allow OTA Devices to receive configuration updates from the OBFCM Data Service, we use the dedicated topic iot4emissions/configServer/<device-ID> on the MQTT broker of the OBFCM Data Service.

Whenever a new configuration update must be sent to a device, the OBFCM Data Service will internally publish a retained message to the above topic for the device concerned. This message will automatically be retrieved by the OTA Device as soon as it connects to the MQTT broker and subscribes to its device-specific topic. This is a point-to-point communication pattern.

Data transfer topic

To allow OTA Devices to send OBFCM data to the OBFCM Data Service, we use the dedicated topic iot4emissions/dataDevice/<device-ID> on the MQTT broker of the OBFCM Data Service. When the time comes for periodic data transfer, each OTA Device will connect to the MQTT broker of the OBFCM Data Service and will publish its data to its device-specific topic.

3.1.2 Topic access rights

OTA Devices should be granted access to the following MQTT topics:

- iot4emissions/registerDevice/<device-ID> (publish right)
- iot4emissions/registerServer/<device-ID> (subscribe right)
- iot4emissions/configServer/<device-ID> (subscribe right)
- iot4emissions/dataDevice/<device-ID> (publish right)

The MQTT access policy for the OTA Registration Service should be:

- iot4emissions/registerDevice/* (subscribe right)
- iot4emissions/registerServer/* (publish right)

Conversely, the MQTT access policy for the OBFCM Data Service should be:

- iot4emissions/configServer/* (publish right)
- iot4emissions/dataDevice/* (subscribe right)

3.2 MQTT message types

Table 3 presents the types of MQTT messages which are exchanged between the OTA Device and the MQTT brokers of the OTA Registration Service and OBFCM Data Service.

Table 3. MQTT message types

ID	Message type	OTA Device state	Sender	Receiver
1	Registration Request	Provisioned	OTA Device	Registration Service
2	Registration Status	Provisioned	Registration Service	OTA Device
3	Configuration Status	Registered, Initialized or Operational	OBFCM Data Service	OTA Device
4	Data Transfer	Operational	OTA Device	OBFCM Data Service

Source: JRC, 2022

3.3 MQTT message structure

3.3.1 Registration Messages

Initially, the OTA Device needs to register itself by sending a request message to the OTA Registration Service. The prerequisite actions for this are a) connecting successfully to the MQTT broker used by the OTA Registration Service, and b) subscribing to the topic iot4emissions/registerServer/<device-ID> where retained registration status messages are published (this retained message lets the OTA Device know that the registration process has been completed successfully).

As already discussed, the topic to which registration request messages are published is iot4emissions/registerDevice/<device-ID>. The message payload is a stringified JSON object containing information from the vehicle's Certificate of Conformity (CoC). This can be a subset of those specified in Regulation (EU) 2019/631 for new passenger cars (Section 2, Part B of Annex II to 2019/631) and for new light commercial vehicles (Section 2, Part C of Annex III to 2019/631). An indicative list of vehicle attributes common to both passenger cars and light commercial vehicles is provided in Table 4 below.

Table 4. Structure of Registration Request message

Field name	Field description	Measurement unit	Remarks
`vin`	Vehicle Identification Number	-	JSON string representation of standardized identifier (36 characters)
`vehicleInfo`	Set of vehicle attributes	-	JSON object
`vfn`	Vehicle family identification number	-	JSON string representation of standardized identifier (24 characters)
`mh`	Name of the manufacturer EU standard denomination	-	JSON string representation of descriptive text
`tan`	Type approval number	-	JSON string representation of descriptive text
`t`	Type	-	JSON string representation of descriptive text
`va`	Variant	-	JSON string representation of descriptive text
`ve`	Version	-	JSON string representation of descriptive text
`mk`	Make	-	JSON string representation of descriptive text
`mt`	WLTP test mass	kg	JSON string representation of integer number
`ewltp`	Specific emissions of CO_2 (WLTP)	g/km	JSON string representation of integer number
`z`	Electric energy consumption	Wh/km	JSON string representation of integer number
`it`	Code for innovative technology or group of innovative technology	-	JSON string representation of descriptive text
`erwltp`	Total WLTP CO_2 emissions reduction due to an innovative technology	g/km	JSON string representation of floating-point number with two decimal places

Source: JRC, 2022

An example Registration Request message is presented in Box 1 below for an actual vehicle type (internal combustion engine) with a fictional VIN number.

Box 1. Registration Request message example

```
{
    "vin": "JFXHF05K1L4104934",
    "vehicleInfo": {
       "vfn": "IP-03_MX_0169-1C4-1",
       "mh": "FIAT GROUP",
       "tan": "E4*2007/46*1410*01",
       "t": "MX",
       "va": "JHVFV",
       "ve": "K5LE1C",
       "mk": "JEEP",
       "mt": "1872",
       "ewltp": "157"
    }
}
```

When the registration process is completed successfully, the OTA Device will receive the following message on the topic iot4emissions/registerServer/<device-ID>. This message is published to the MQTT broker as a retained message, which means that if the OTA Device is not connected when the message is published, the message will remain stored on the broker to be read the next time that the OTA Device will connect. Unlike normal MQTT messages which get discarded by the MQTT broker if there is no device subscribed to receive them, retained messages can persist on the broker until they are read or until a certain time period has lapsed. This is configured by setting the "retained message" flag to true.

Table 5. Structure of retained Registration Status message

Field name	Field description	Remarks
deviceState	Device state name (registered or error)	JSON string representation of device state or error message

Source: JRC, 2022

An example of a retained Registration Status message is presented in Box 2 below.

Box 2. Registration Status message example

```
{
    "deviceState": "registered"
}
```

3.3.2 Configuration/Initialization Messages

After an OTA Device has been registered it checks periodically with the OBFCM Data Service for new configuration updates. This is done by connecting to the MQTT broker of the OBFCM Data Service and subscribing

to the topic iot4emissions/configServer/<device-ID>. The latest configuration update applicable to the specific OTA device will be made available to the device as soon as it connects. As mentioned earlier, these configuration status messages have been previously published to the OBFCM Data Service MQTT broker as retained messages.

The retained configuration status message includes information about:

- the version number of this configuration,
- the network addresses of the MQTT brokers that the OTA Device should be connecting with,
- the frequencies of certain device actions such as acquiring data, transferring data and checking for new configuration updates.

The retained message also includes a command to the OTA Device about its desired functional state, i.e., whether it should be in an initialized or operational state. This state transition command will depend on whether or not the OTA Device has previously been added to the "allow-list" of the OBFCM Data Service, which means it should be collecting & periodically sending OBFCM data.

Table 6. Structure of retained Configuration Status message

Field name	Field description	Remarks
`deviceState`	Device state name	JSON string representation of device state
`configuration`	Set of configuration parameters	JSON object containing connection and frequency configuration information
`registrationMqtt`	Set of connection parameters for the MQTT broker of the Registration Service	JSON object containing endpoint, port and protocol information
`obfcmDataMqtt`	Set of connection parameters for the MQTT broker of the OBFCM Data Service	JSON object containing endpoint, port and protocol information
`endpoint`	Network address	JSON string representation of internet host address
`port`	Network port number	JSON string representation of communication port
`protocol`	Message transport protocol	JSON string representation of message protocol
`acquireFreq`	Frequency at which the OTA Device will be acquiring values from the OBFCM device	JSON string representation of a cron schedule expression
`transferFreq`	Frequency at which the OTA Device will be transferring OBFCM values to the Commission	JSON string representation of a cron schedule expression
`configCheckFreq`	Frequency at which the OTA Device will be checking for new configuration updates	JSON string representation of a cron schedule expression
`version`	Version number of the configuration settings	JSON string representation of configuration version identifier

Source: JRC, 2022

An example of a retained Configuration Status message is presented in Box 3 below, instructing the OTA Device to transition, acquire data from the OBFCM device once every month, and transfer this data every three months.

Box 3. Configuration Status message example


```
{
	"deviceState": "operational",
	"configuration": {
		"registrationMqtt": {
			"endpoint": "a1u4qvcqj4uf51-ats.iot.eu-central-1.amazonaws.com",
			"port": 8883,
			"protocol": "mqtts"
		},
		"obfcmDataMqtt": {
			"endpoint": "a1u4qvcqj4uf51-ats.iot.eu-west-1.amazonaws.com",
			"port": 8883,
			"protocol": "mqtts"
		},
		"acquireFrequency": "0 0 0 ? * * *",
		"transferFrequency": "0 0 0 ? 1/3 * *",
		"configCheckFrequency": "0 0 0 ? 1/3 * *",
		"version": 1
	}
}
```


3.3.3 Data Transfer Messages

When the OTA Device switches to an operational state, it periodically collects OBFCM data from the vehicle's OBFCM device and sends them to the OBFCM Data server. This is done by publishing to the topic iot4emissions/dataDevice/<device-ID>. The "data" field of the message is a JSON array containing objects that feature several fields as illustrated in Table 7 below.

An example of a Data Transfer message is given in Box 4 below, containing three snapshots of the OBFCM device values for the total fuel consumed and total distance travelled parameters. The values have been acquired at three different points in time (June 1st, July 1st, August 1st 2021) and are being sent in a single data transfer operation.

Table 7. Structure of Data Transfer message

Field name	Field description	Measurement unit	Remarks
data	Set of data values acquired from the OBFCM device	-	JSON array
acquired	UNIX timestamp representing the point in time when the data was acquired	sec	JSON string representation of UNIX timestamp. Max character length: 10
tfc	Total fuel consumed (lifetime)	lt	JSON string representation of floating-point number with one decimal place
tdt	Total distance travelled (lifetime)	km	
tfccdo	Total fuel consumed in charge depleting operation (lifetime)	lt	
tfcdscio	Total fuel consumed in driver-selectable charge increasing operation (lifetime)	lt	
tdtcdoeo	Total distance travelled in charge depleting operation with engine off (lifetime)	km	
tdtcdoer	Total distance travelled in charge depleting operation with engine running (lifetime)	km	
tdtdscio	Total distance travelled in driver-selectable charge increasing operation (lifetime)	km	
tgeib	Total grid energy into the battery (lifetime)	kWh	

Source: JRC, 2022

Box 4. Data Transfer message example

```
{
    "data": [
        {
            "acquired": 1622538000,
            "tfc": 32.5,
            "tdt": 295.9,
        },
        {
            "acquired": 1625130000,
            "tfc": 37.9,
            "tdt": 345.2,
        },
        {
            "acquired": 1627808400,
            "tfc": 43.3,
            "tdt": 394.5,
        }
    ]
}
```

4 Proof of concept implementation

A Proof-of-Concept (PoC) implementation was developed to validate the communication protocol's applicability and derive fine-grained technical implementation requirements.

A PoC is also a necessary step in order to produce estimates on the cost of developing and operating the OTA data collection network.

The first iteration of our PoC implementation included the following:

- Development of the internal logic and external interfaces of the **three fundamental technology components** as previously defined in the solution architecture framework, i.e., the OTA Device, the cloud-based OTA Registration Service and the cloud-based OBFCM Data Service.
- Deployment of two cloud-based **MQTT broker services** (Amazon AWS IoT Core) to mediate the communication between the OTA Device and the two cloud services. One MQTT broker service mediates communication between the OTA Devices and the OTA Registration Service (AWS data centre in Frankfurt). The other MQTT broker is between OTA Devices and the OBFCM Data Service (AWS data centre in Ireland).
- Conceptualization of a **Public Key Infrastructure management procedure** for introducing trusted producers of OTA Devices into the trust chain of the OTA data collection network. This is done by registering the X.509 certificate of every OTA Device producer with the AWS IoT Core service, assigning trusted Certificate Authority status to the OTA Device producer.
- Implementation of a procedure for **Just-In-Time Provisioning** of security certificates to OTA Devices (upon their first connection to AWS) to enable secure encrypted communication between OTA Devices and cloud systems.
- Development of a **web-based application for small-scale demonstration**, which showcases the real-time interactions and operation of the entire OBFCM data collection network through a web browser.
- Development of a **Command Line Interface (CLI)-based large-scale fleet simulation system** that can instantiate large sets of vehicles with configurable “behaviour” parameters (e.g., annual mileage rate, fuel consumption rate, frequency of data transfers and configuration requests). The simulation system provides the basis to measure the exact data volume and number of message interactions that are generated with our off-board communication protocol. This is essential to inform estimations regarding the operating cost of the OTA data collection network.
- Development of a system that stores the OBFCM data received from simulated OTA Devices in a **time-series database** for post-processing analytics and provides **visualization dashboards**.

In a second iteration of the PoC we expanded the architecture and evolved the implementation to enable operation at a large scale and support fleet-wide data collection. This architecture is described in Chapter 5.

4.1 Systems developed

4.1.1 OTA Registration Service application

Functionality

The OTA Registration Service maintains a database of registered OTA Device IDs mapped to the VINs of the vehicles on which those devices are installed. It also stores basic information about the vehicle’s technical characteristics (manufacturer, model, engine type etc). The OTA Registration Service receives MQTT registration requests from OTA Devices, stores the above-mentioned information about the device and the vehicle, notifies the OBFCM Data Service of the newly registered device by HTTP API, and confirms the registration back to the OTA Device via MQTT retained messages.

The OTA Registration Service also has an HTTP API through which it receives notifications by a Vehicle Registrar organization (this could be the vehicle manufacturer or a Member State transport authority). These notifications arrive when a vehicle is sold and registered in an EU Member State. Their purpose is to let the OTA Registration

Service know that this vehicle is cleared to collect and report OBFCM data. This information is stored by the OTA Registration Service and, as already mentioned, is forwarded to the OBFCM Data Service via an HTTP API call.

Implementation

In the first, small-scale, iteration of our PoC implementation, the server application listened for messages from OTA Devices through an MQTT client connected to an MQTT broker provided by AWS IoT Core. The server application also sent confirmation messages to the devices through the same MQTT client. Through its MQTT client, the Device Listener component also handled all MQTT related events such as connections, errors, subscriptions, incoming and outgoing data. In the second iteration of the PoC which supports large scale operation, all interactions between the OTA Registration Service application and the MQTT broker take place via scalable AWS Lambda functions. More information is provided in Chapter 5.

The OTA Registration Service also features a client component for the HTTP API exposed by the OBFCM Data Service. This client communicates with the OBFCM Data Service via web services to notify the OBFCM Data server about new OTA Device registrations and activations. An OTA Registry component handles information about OTA Devices and a respective Vehicle Registry component handles information about vehicles.

Table 8. OTA Registration Service implementation components

Component	Type	Description
OTA Registration Service application	application	Node.js v12.22.1 application, exposes web services and connects to MQTT broker
MongoDB	database	MongoDB v4.4.7
Node.js	javascript runtime	Node.js v12.22.1 runtime environment
Docker	Container service	Server Version v20.10.8, Docker Compose v2.0.0-rc.2
Ubuntu	OS	Ubuntu v20.04.2, hosts our applications and third-party software
AWS IoT Core	IoT services	Provides secure MQTT brokers and Just-In-Time Provisioning (JITP) for IoT devices

Source: JRC, 2022

4.1.2 OBFCM Data Service application

Functionality

The OBFCM Data Service application is responsible for a) accepting and handling OTA Device registration & vehicle registration notifications from the OTA Registration Service, b) publishing configuration updates, and c) receiving and storing OBFCM data that OTA Devices send.

The OBFCM Data Service receives notifications from the OTA Registration Service whenever a new OTA Device is registered and later when the vehicle enters circulation and the device is cleared to start transmitting OBFCM data. The server also instructs the devices to start/stop collecting and transmitting OBFCM data through configuration updates.

Implementation

To enable communication with the OTA Registration Service several web services are exposed by the OBFCM Data Service application. Similarly to the previously described implementation of the OTA Registration Service application, the communication of this application with the OTA Devices was handled by a Device Listener component. In the second iteration, this was replaced by communication via AWS Lambda functions.

The OBFCM Data Service also features an OTA Registry component to keep track of the OTA Devices of the system and handle device registration and activation. Finally, the OBFCM Data Service also connects to a time series optimized database to store the incoming OBFCM value snapshots per vehicle (lifetime distance travelled, lifetime fuel consumed, etc). Early in our PoC implementation process, we used InfluxDB and later QuestDB. In the second iteration, this was replaced by AWS Timestream.

Table 9. OBFCM Data Service implementation components

Component	Type	Description
OBFCM Data Service application	application	Node.js v12.22.1 application exposes web services and connects to MQTT broker
InfluxDB	database	InfluxDB v1.8.6, stores OBFCM data (time series)
QuestDB	database	QuestDB v6.0.4, stores OBFCM data (time series)
MongoDB	database	MongoDB v4.4.7
Node.js	javascript runtime	Node.js v12.22.1 runtime environment
Docker	Container service	Server Version v20.10.8, Docker Compose v2.0.0-rc.2
Ubuntu	OS	Ubuntu v20.04.2, hosts our applications and third-party software
AWS IoT Core	IoT services	Provides secure MQTT brokers and Just-In-Time-Provisioning for IoT devices

Source: JRC, 2022

4.1.3 Fleet simulator application

Functionality

The Fleet Simulator is an application that simulates one or more vehicles equipped with an OTA Device. The OTA Device of each simulated vehicle starts operating when the vehicle is virtually powered on. The OTA Device communicates with the vehicle's on-board systems via direct function calls, since software modules simulate the on-board systems.

On the other hand, communication with off-board systems is not virtual, but actual. The OTA Device communicates with the OTA Registration Service and the OBFCM Data Service exclusively via secure MQTT. The MQTT client used in an OTA Device requires the MQTT broker's endpoint information and a set of credentials to connect securely using the mqtts protocol and perform mutual authentication. The MQTT broker's address, port and protocol are stored in non-volatile memory as factory-default settings of the OTA Device.

The public full-chain certificate of the device is required and the associated private key to encrypt communication and verify the device's identity to the broker. The root certificate of the MQTT broker is also required to decrypt incoming communication and verify the identity of the broker. After successful authentication, a connection to the MQTT broker has been established and the OTA Device may publish messages or subscribe to topics according to the access rights we have stated in a provisioning template. Our implementation uses MQTT Quality of Service (QoS) level 1. This level guarantees that messages sent from OTA Devices to the MQTT broker will be delivered at least once, without excluding the possibility to be delivered more than once.

Before the OTA Device can acquire and transfer OBFCM value readings it must send a registration request to the OTA Registration Service and wait for confirmation. After the device registration has been confirmed, the OTA Device connects to the MQTT broker of the OBFCM Data Service and automatically receives its initial configuration settings (as retained message). Eventually, the OTA Device will receive an instruction from the OBFCM Data Service to start transmitting OBFCM readings.

The Fleet Simulator application coordinates the behaviour of different OTA Devices according to a preconfigured vehicle "simulation profile". This profile can be adapted to control parameters such as the average annual mileage of each vehicle under simulation, its average fuel consumption rate and its connectivity rate. These parameters allow executing the OTA data collection network simulations with different scenarios.

Implementation

The Fleet Simulator application can spawn several simulated vehicle instances. The application will parse a folder where vehicle data and simulation profiles are stored at startup. It will spawn the respective vehicle instances and instruct them to start operating.

A simulated vehicle is a class that has

1. a CAN Bus component that provides static information such as VIN and vehicle model info,
2. an OBFCM device component that provides OBFCM data such as total distance travelled and total fuel consumed (according to some simulation profile), and
3. an OTA Device component that acquires information from the CAN Bus component to register itself with the OTA Registration Service, and then acquires data from the OBFCM device component in order to transmit it to the OBFCM Data Service.

The OTA Device component is a class that has:

1. a simulated Trusted Platform Module (TPM) component responsible for holding credentials such as private keys and X.509 certificates,
2. a device logger to log important events and errors,
3. two MQTT client instances to communicate with the OTA Registration Service's MQTT broker and the OBFCM Data Service's MQTT broker.

The OTA Device component holds the OTA Device's state and is responsible for transitioning to another state or transitioning to its previous state after a device reboot occurs. It is responsible for acquiring data from the OBFCM device, storing them to the local storage and transmitting them to the OBFCM Data server, while in operational mode. Through its MQTT clients, the OTA Device component connects to the MQTT brokers and handles all MQTT related events such as connections, errors, subscriptions, incoming and outgoing data.

To perform repeated tasks at regular intervals such as requesting registration / configuration data, the OTA Device component uses a cron-like scheduler.

Error conditions which are due to network issues (e.g., loss of mobile network connectivity, low network bandwidth or high latency) or due to server issues (e.g., dropped / half-closed TLS network connection, MQTT broker connection throttling, internal server failure) are handled by the OTA Device through an automatic reconnection procedure that implements an exponential back-off algorithm[3].

Table 10. Fleet Simulator implementation components

Component	Type	Description
Fleet Simulator	application	Node.js v12.22.1 application exposes web services and connects to MQTT broker
MongoDB	database	v4.4.7
Node.js	javascript runtime	Node.js v12.22.1 utilized by our applications
Docker	Container service	Server Version v20.10.8, Docker Compose v2.0.0-rc.2
Ubuntu	OS	Ubuntu v20.04.2, hosts our applications and third-party software

Source: JRC, 2022

4.1.4 Certificate Authority simulator application

Functionality

In a future full-scale deployment of the off-board communication solution presented in this report, the digital certificates of OTA Devices will be created and signed by those organizations who will be producing the OTA Devices, i.e., by Tier 1 component suppliers or vehicle manufacturers. For the purposes of our Proof-of-Concept

[3] https://aws.amazon.com/builders-library/timeouts-retries-and-backoff-with-jitter/

implementation we needed to be able to simulate this procedure of issuing certificates for large batches of OTA Devices.

For this purpose, we developed an application that simulates the operation of a Certificate Authority (CA). The CA simulator accepts Certificate Signing Requests (CSRs) and issues signed certificates for newly created OTA Devices. The digital certificate which testifies the identity of the CA is self-signed. This CA simulator allows us to quickly generate X.509 certificates for large numbers of OTA Devices that we store and reuse across different fleet simulation runs.

A basic feature of the CSR procedure implemented by our CA simulator is that the OTA Device ID is placed in the Common Name field of the device's certificate. This enables the AWS IoT service to later read the OTA Device ID from the device's certificate and use the ID in applying an MQTT topic access policy and restricting each device's access to device-specific topics only.

Implementation

Before an OTA Device is able to connect securely to the MQTT broker, the Certificate Authority (CA) that has issued the device's certificate must have already been registered with the AWS IoT. It's not enough to simply upload the self-signed certificate of the Certificate Authority to the AWS IoT service. The registration process requires proof that whoever appears to represent the Certificate Authority (CA) holds the private key associated with the CA certificate. This is accomplished by creating a private key verification certificate.

The process comprises of the following steps (using openssl):

1. Generate a key pair for the private key verification certificate,
2. Create a Certificate Signing Request (CSR) where the Common Name field has the value of a unique code provided by the AWS IoT,
3. Use the CSR to create the verification certificate signed by your CA, and
4. upload the CA certificate and the verification certificate to the AWS IoT.

This process must be done in real time with the cooperation of the AWS infrastructure administrator since the unique code generated by the AWS, which is involved in the process, is not guaranteed to be always the same.

Table 11. Certificate Authority Simulator implementation components

Component	Type	Description
CA Simulator	application	Node.js v12.22.1 application, exposes web service to create private key and signed X.509 certificate
Node.js	javascript runtime	Node.js v12.22.1 utilized by our applications
Docker	Container service	Server Version v20.10.8, Docker Compose v2.0.0-rc.2
Ubuntu	OS	Ubuntu v20.04.2, hosts our applications and third-party software

Source: JRC, 2022

4.1.5 Workflow demonstrator application

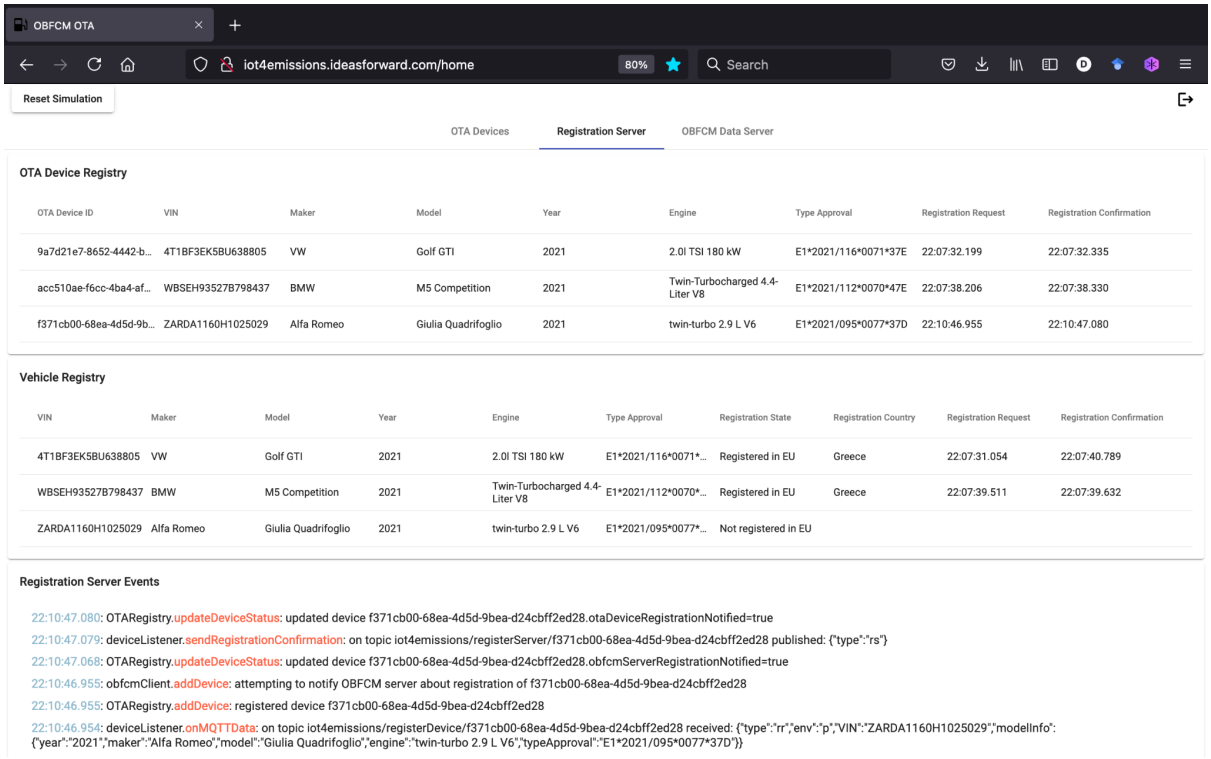
Functionality

This is a simple web application built to offer an interactive visual demonstration of the entire PoC workflow. It offers three different views into events taking place on the OTA data collection network and on how the PoC components interact with each other.

OTA Registration Service view:

- A view on the content and updates happening on the OTA Registration Service's device registry and vehicle registry
- Real-time feed of MQTT publish and subscribe events taking place on the MQTT broker of the OTA Registration Service.

Figure 5: Workflow demonstrator - OTA Registration Service view

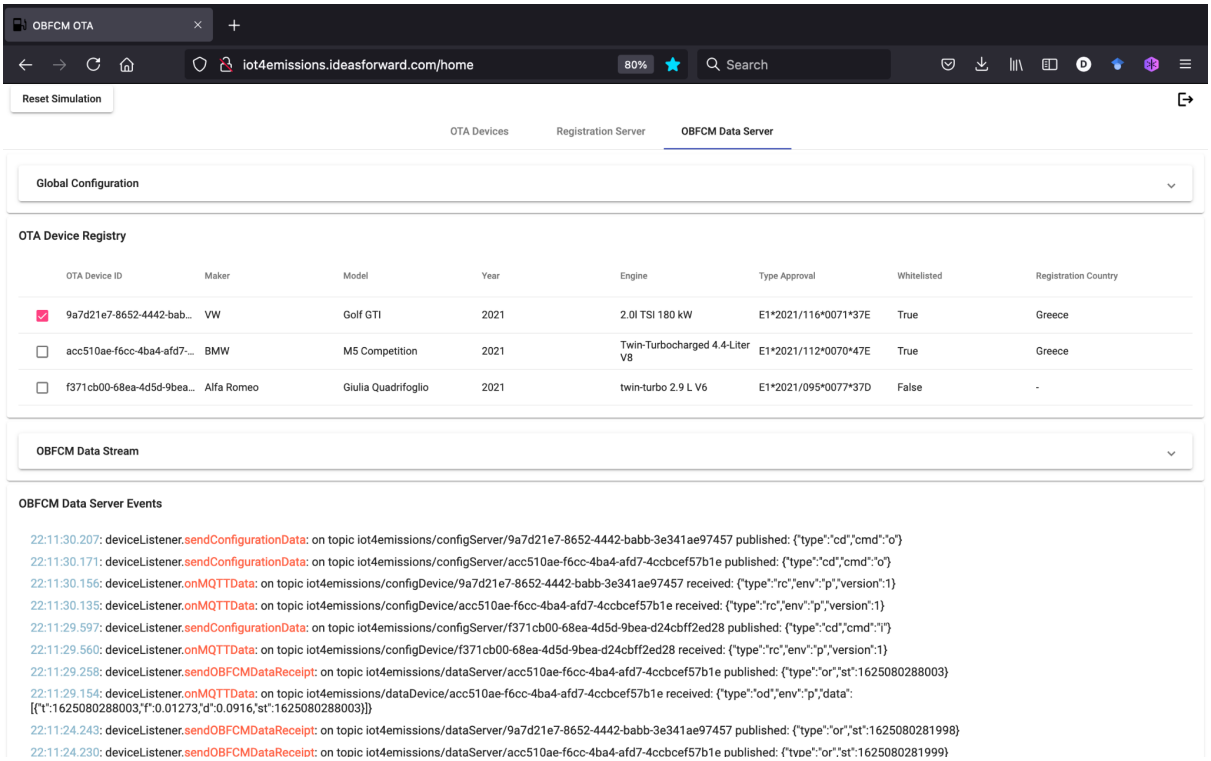


Source JRC 2022

OBFCM Server view:

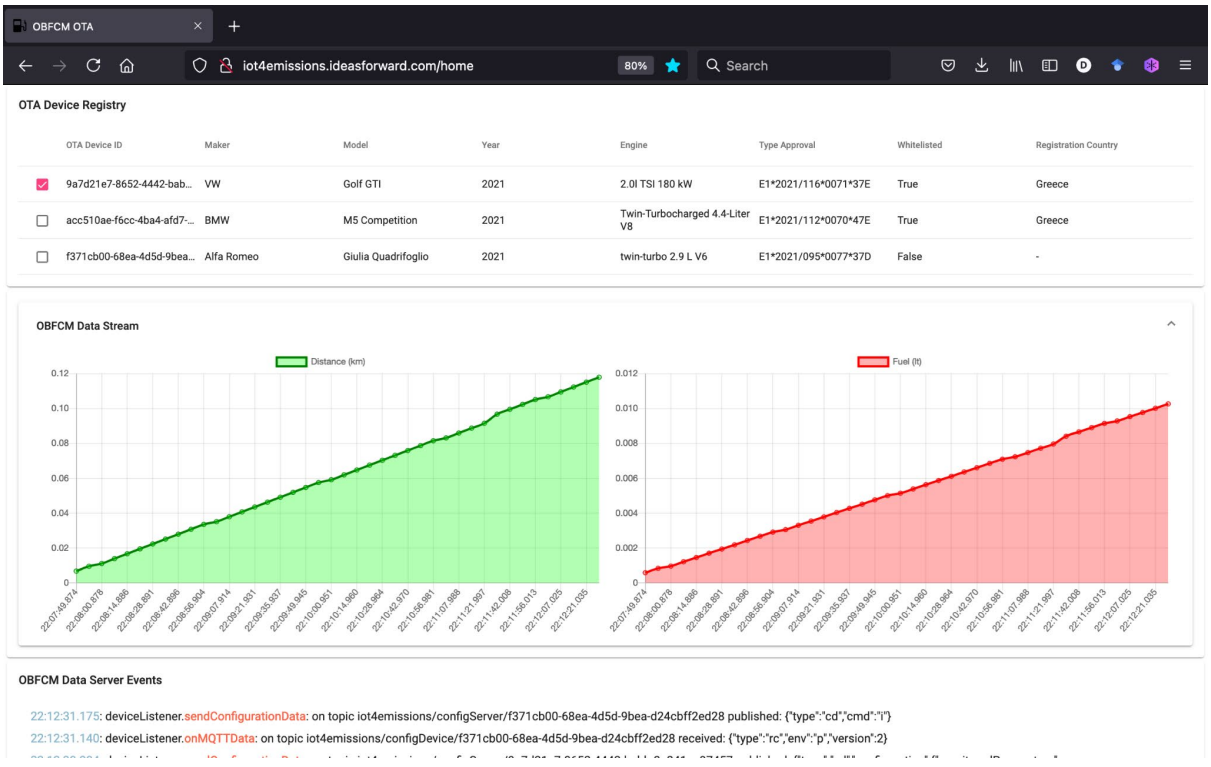
- View on the content and updates on the OBFCM Data Service's device registry;
- Real-time feed of MQTT publish and subscribe events taking place on the MQTT broker of the OBFCM Data Service;
- View of the incoming OBFCM data streams from each vehicle (lifetime distance and lifetime fuel values over time).
- Ability to modify the global configuration parameters (data acquisition and data transfer frequency) in order to see how the change is propagated to devices and how it affects their behaviour.

Figure 6: Workflow demonstrator - OBFCM Data Service view



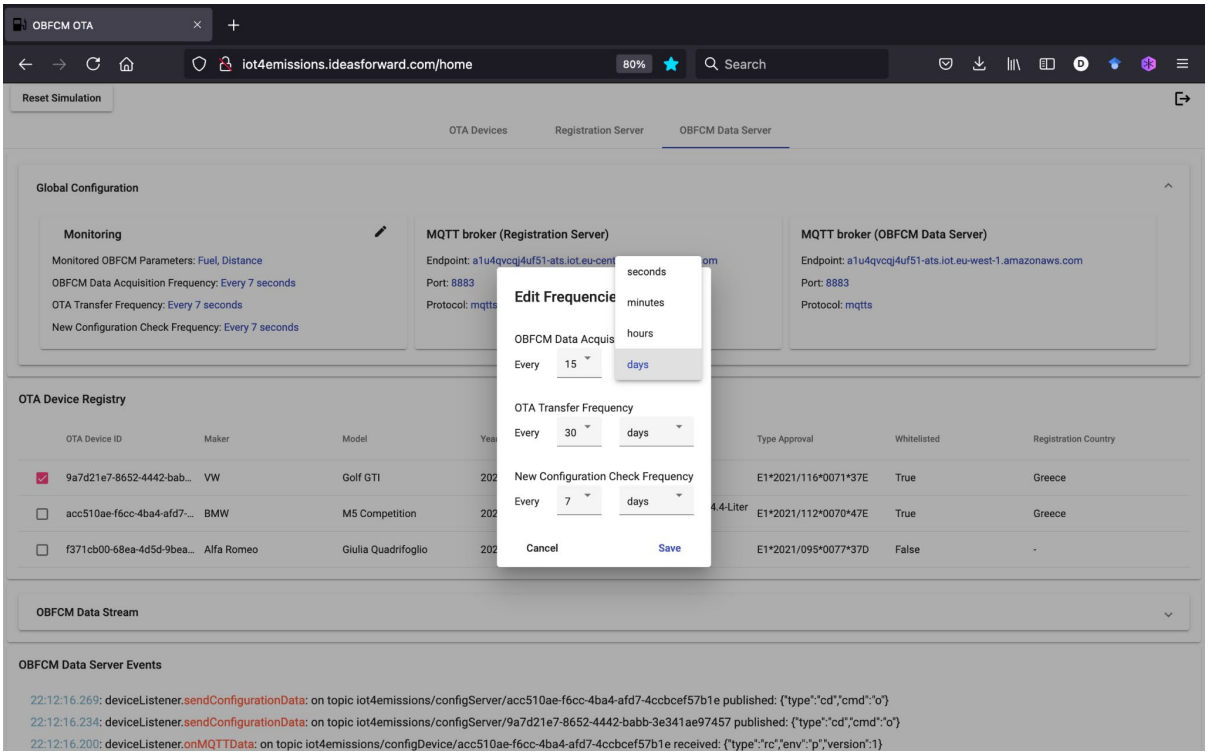
Source JRC 2022

Figure 7: Workflow demonstrator –streams of OBFCM parameter values from a vehicle



Source JRC 2022

Figure 8: Workflow demonstrator - modifying the global OTA Device configuration settings



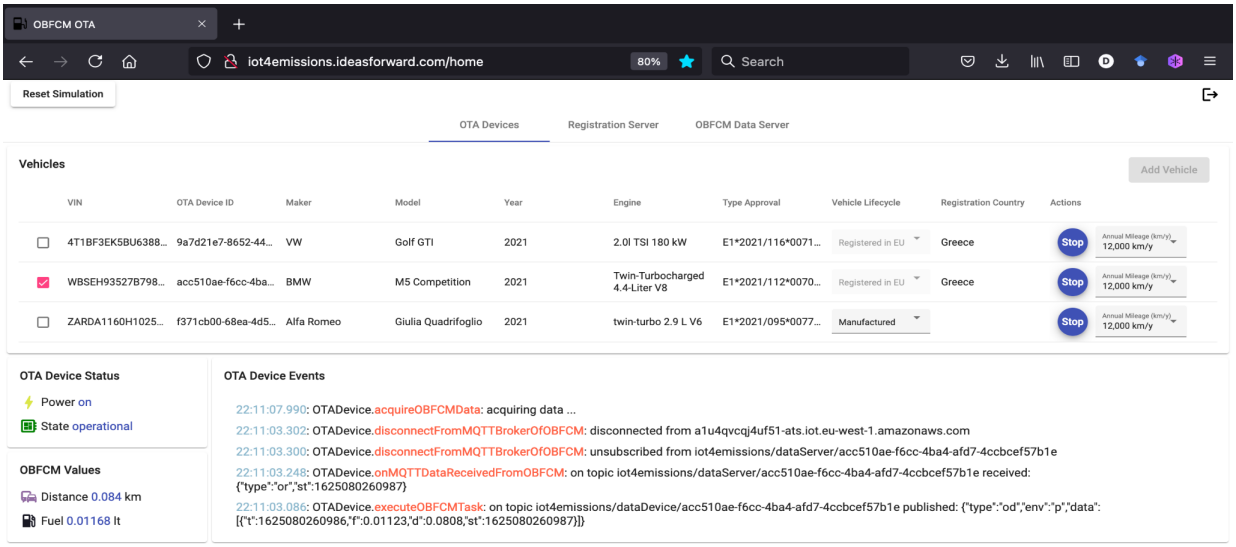
Source JRC 2022

Fleet Simulator / OTA Device view:

- Ability to populate the demonstration fleet with a small number of vehicles each of which has its preset profile data (VIN, OTA Device ID, manufacturer, model, year, engine type, type approval number);
- Ability to start and stop the engine of each vehicle, triggering the OTA Device's first MQTT connection to the OTA Registration Service;
- Ability to post an update to the OTA Registration Service to simulate the event of a vehicle's registration in an EU Member State;
- Ability to modify part of each vehicle's simulation profile by changing its annual mileage setting.
- View of the incoming OBFCM data streams from each vehicle (lifetime distance and lifetime fuel values over time) on an external visualization dashboard powered by Grafana, displaying real-time updates from a back-end time series database.

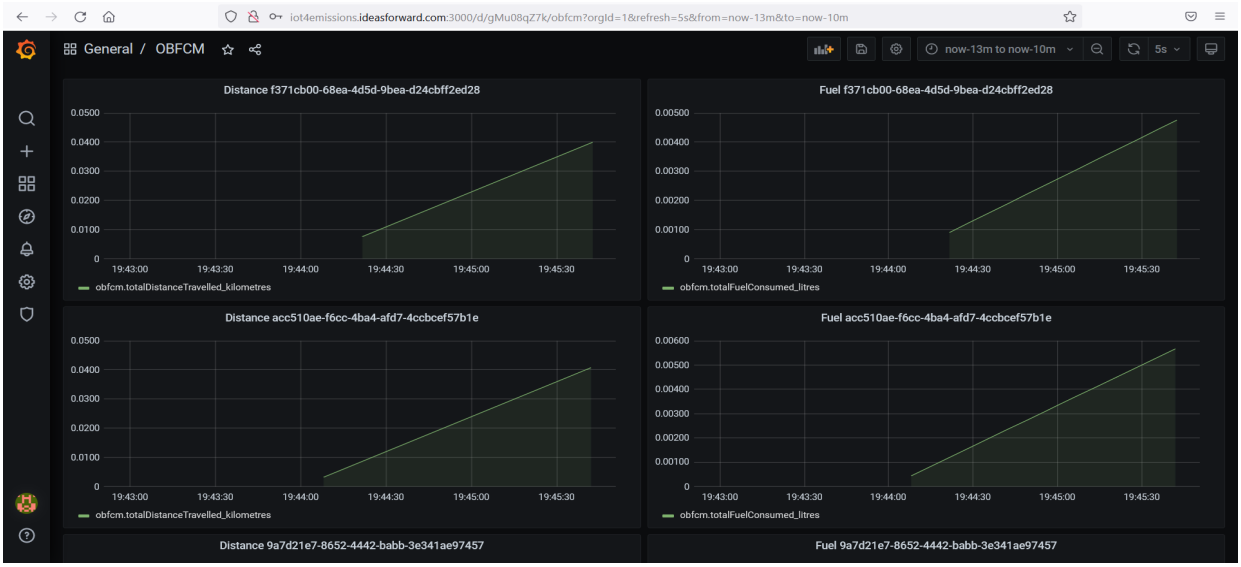
The workflow demonstration app is complemented by a data visualization dashboard created with Grafana. The dashboard is shown in Figure 10, displaying lifetime distance and fuel for two vehicles.

Figure 9: Workflow demonstrator - Home dashboard



Source JRC 2022

Figure 10: External OBFCM data visualization dashboard



Source JRC 2022

Implementation

Table 12. Workflow demonstrator implementation components

Component	**Type**	**Description**
DemoApp	Application	Angular v11 web application, consumes web services
Nginx	Web server	nginx v1.18.0, serves web application
Grafana	Visualization toolkit	Grafana v8.0.2, visualizes data stored in the InfluxDB
Docker	Container service	Server Version v20.10.8, Docker Compose v2.0.0-rc.2
Ubuntu	OS	Ubuntu v20.04.2, hosts our applications and third-party software

Source: JRC, 2022

4.2 Processes implemented

Subsections 4.2.1 to 4.2.3 describe the three most important processes in which the OTA Device, the OTA Registration Service and the OBFCM Data Service are engaged, as implemented in our Proof-of-Concept.

The order in which the processes are presented corresponds to the actual progression of evens in the lifecycle of a vehicle.

First, as soon as the vehicle's production is completed, the OTA Device communicates with the OTA Registration Service for the first time in order to register. After successful authentication and registration, the device subsequently communicates with the OBFCM Data Service for the first time to receive its initial configuration settings. At this point, it is put into waiting mode.

Second, as soon as the OTA Registration Service receives knowledge that the vehicle has been registered in an EU country and that any legal preconditions to data collection are now satisfied, an intra-service communication takes place to inform the OBFCM Data Service.

Third, on the next occasion where the OTA Device communicates with the OBFCM Data Service to check for updates to its configuration settings, it receives instructions to start the periodic transfer of OBFCM data. At this point the OTA Device is in full operational mode.

4.2.1 OTA Device becomes registered and initialized

Figure 11 below presents a UML sequence diagram with the interactions taking place between OTA Device, Registration Service and OBFCM Data Service as soon as a new vehicle has been produced.

The step-by-step description provided in the text after the figure is based on our current implementation of the respective process.

4.2.1.1 Registration

OTA Device wakes up

If this is the first time that the OTA Device is put in operation it will start with acquiring the resources necessary from its local environment to initiate communication with the OTA Registration Service. Specifically,: a) it acquires the necessary credentials (private key, certificates) from the simulated Trusted Platform Module (TPM), b) it reads the VIN and vehicle model information from the simulated CAN Bus and c) it loads from local storage the factory configuration that contains the addresses of the two MQTT brokers that enable communication with the outside world.

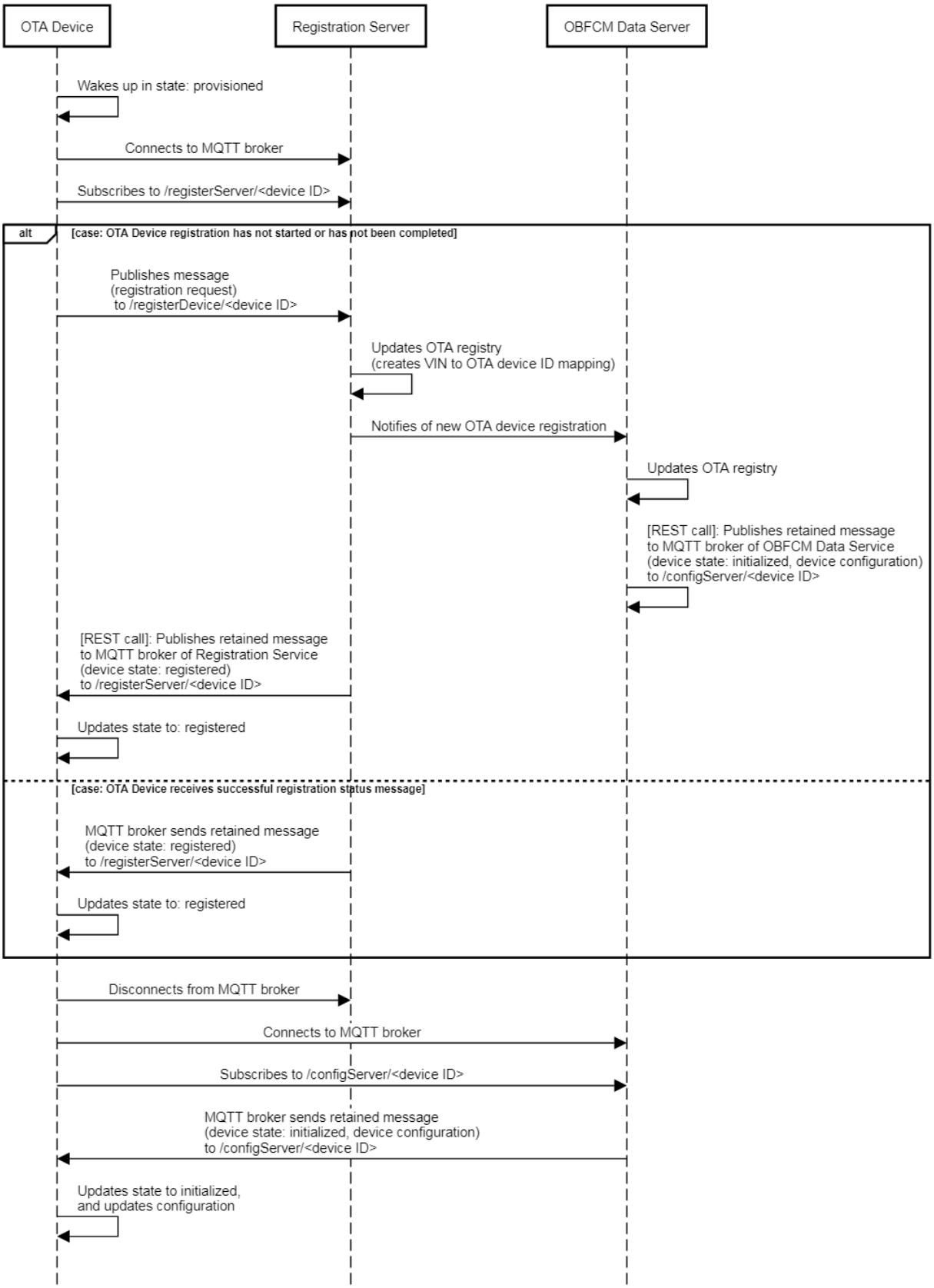
OTA Device connects to OTA Registration Service

Afterwards, the OTA Device attempts to connect to the MQTT broker of the OTA Registration Service. In our implementation, the MQTT broker of the OTA Registration Service is a managed service provided by AWS IoT Core. To facilitate mutual authentication between the systems on the two sides, the OTA Device is provisioned with the AWS Root certificate and the AWS IoT Core service is conversely provisioned with the certificate of the Certificate Authority that signed the OTA Device's certificate. Connecting to the MQTT broker triggers a process called Just-In-Time-Provisioning which automates authentication and management of access rights.

OTA Device subscribes for registration status messages

Upon successful completion of the Just-In-Time-Provisioning process, the OTA Device subscribes to the topic where the OTA Registration Service publishes registration status messages for the specific device. For obvious security purposes, an OTA Device is only allowed to publish and subscribe to MQTT topics destined exclusively for that device.

Figure 11: OTA Device becomes registered and initialized



Source JRC 2022

4.2.1.2 Case A: OTA Device registration has not started or has not been completed

At this point in time there are two alternative execution paths depending on the case at hand. If this is the first time that the OTA Device has connected to the Registration Service then there will be no retained Registration Status message waiting to be delivered to the OTA Device.

OTA Device requests registration

The OTA Device sends a registration request to the OTA Registration Service by publishing to the respective MQTT topic. The request contains the vehicle's VIN and technical info. It then waits for a successful registration confirmation (retained message).

OTA Registration Service is updated

The OTA Registration Service handles the registration requests by the OTA Devices. Upon receiving such a request, the server updates its OTA device registry, i.e., a database where the OTA Device IDs and the associated VINs are kept.

Registration Service sends intra-service notification to OBFCM Data Service

Subsequently, the OTA Registration Service notifies the OBFCM Data Service via web service call to handle the device registration on its side.

OBFCM Data Service is updated

The OBFCM Data Service handles the device registration notifications sent by the OTA Registration Service through a web service call (HTTP). When such a notification arrives the OBFCM Data Service adds the OTA Device to its OTA Device registry and marks it as "deny-listed". With the device added to the device registry, the server will now respond to incoming OBFCM data transfer messages from the OTA Device instead of ignoring them. The device registry of the OBFCM Data Service holds limited information about the vehicle to which the OTA Device belongs. The OBFCM Server receives the vehicle model info and the device ID from the OTA Registration Service but it does not receive the Vehicle Identity Number. This is a form of separation of duties based on the principle of least privilege (PoLP). It provides data segregation which elevates data privacy protection.

OBFCM Data Service prepares device configuration

A message is published on the MQTT broker of the OBFCM Data Service, on the device's dedicated MQTT topic for configuration status messages. The message will be retained by the MQTT broker indefinitely (until it is replaced by a new one) and will be automatically delivered to the OTA Device as soon as it connects and subscribes to the respective topic.

OTA Registration Service sends registration status to OTA Device

When a successful response arrives from the OBFCM Data Service to the OTA Registration Service, a message is published on the OTA Registration Service's MQTT broker, on the device's dedicated MQTT topic for registration status messages. The MQTT message is automatically delivered to the OTA Device which updates its internal state flag to "registered" and disconnects from the MQTT broker of the OTA Registration Service.

4.2.1.3 Case B: OTA Device receives successful registration status message

If this is not the OTA Device's first communication with the Registration Service, which means that the device has successfully or unsuccessfully attempted to register in the past, then a retained Registration Status message is automatically delivered to the OTA Device as soon as it subscribes again to the respective topic. The retained message will either confirm registration or notify of an error condition, respectively. For the purposes of this walkthrough, we assume that the registration was successful and that the OTA Device updates its internal state flag to "registered" and disconnects from the MQTT broker of the OTA Registration Service.

4.2.1.4 Initialization

OTA Device connects to OTA Registration Service and requests configuration

Subsequently the OTA Device starts the process in order to move to the “initialized'' state. It connects to the MQTT broker of the OBFCM Data server and subscribes to the device-dedicated topic where the OBFCM Data server publishes configuration updates.

OTA Device receives configuration

As soon as the OTA Device connects to the MQTT broker of the OBFCM Data Service it receives a retained configuration status message. The message contains settings such as up-to-date network connection settings, data acquisition frequency, data transfer frequency, etc. It also includes an instruction to transition to “initialized'' state.

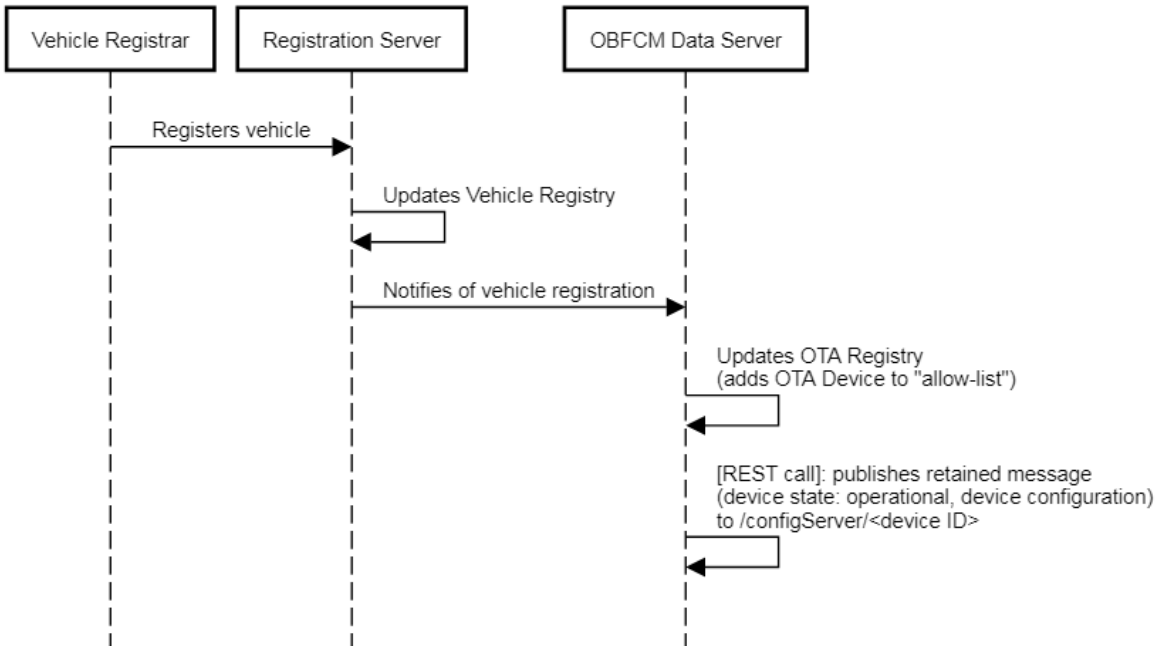
OTA Device updates itself

Every time the OTA Device receives configuration information, it compares it to its current configuration. If it differs, it saves the new configuration in its local storage and reboots in order for the changes to take effect.

4.2.2 OBFCM Server adds OTA Device to allow-list

Figure 12 presents the activities taking place as soon as a new vehicle has been registered in an EU Member State and the Registration Service is updated from an external Vehicle Registrar system (e.g., a national registration authority or vehicle manufacturer).

Figure 12: OBFCM Server adds OTA Device to allow-list



Source JRC 2022

Vehicle Registrar notifies OTA Registration Service

When a new vehicle is sold and is about to enter circulation in an EU Member State, the OTA Registration Service is notified by a remote system (Vehicle Registrar) via a web service (HTTP API) call. The data sent to the OTA Registration Service includes the VIN, the vehicle’s basic technical information and the country of registration. In addition to the requirement that the vehicle is registered in an EU member state, another prerequisite could be that the vehicle owner has granted consent to data collection for the purposes of real-world emissions monitoring by the EC.

OTA Registration Service sends intra-service notification to OBFCM Data Service

The OTA Registration Service updates its Vehicle Registry with the newly registered vehicle. Then, it searches its OTA Registry for the OTA Device ID associated with this vehicle's VIN. If an OTA Device ID is found then the OBFCM Data Service is notified via a web service call that this OTA Device is cleared to collect and transmit OBFCM data.

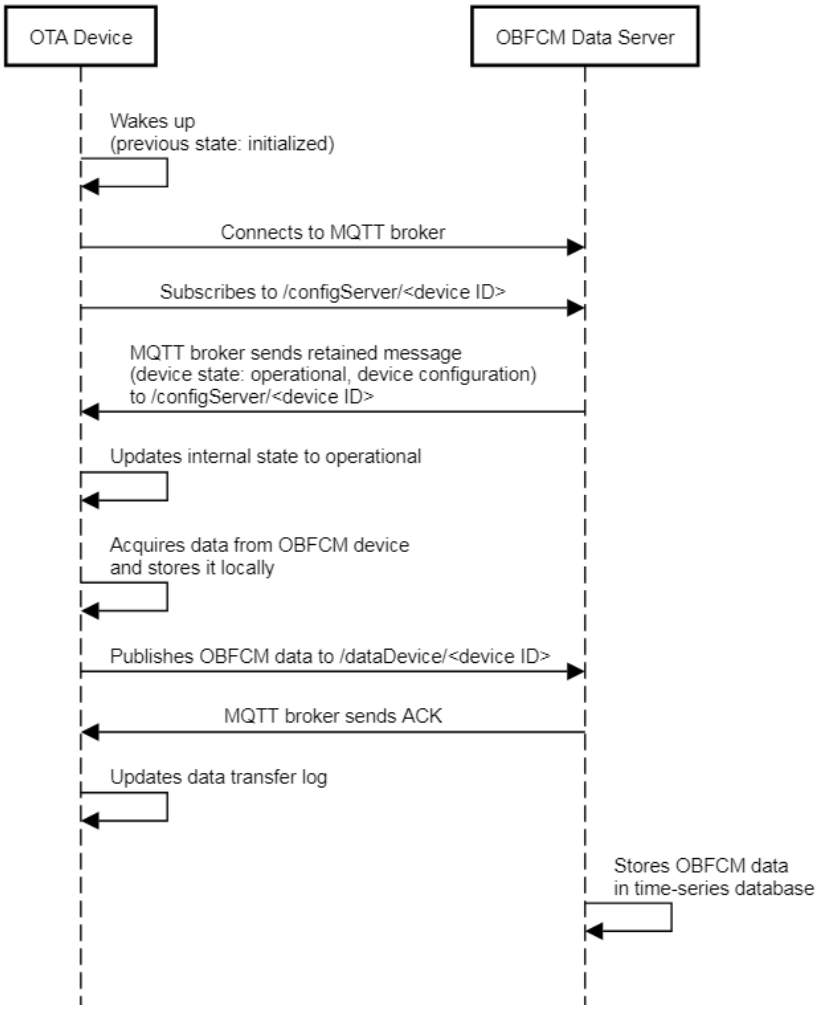
OBFCM Data Service updates its OTA Device registry

When the OBFCM Data server receives this information from the OTA Registration Service it stores it in its OTA Registry. It moves the OTA Device off the data transfer deny-list and onto the allow-list. The OBFCM Data server then publishes a configuration update message on the device's dedicated configuration topic, with an instruction for the OTA Device to move to "operational" state. If the OTA Device is connected to the MQTT broker at that time, it will receive the instruction immediately else it will receive it the next time it connects.

4.2.3 OTA Device becomes operational

Figure 13 shows the sequence of interactions between the OTA Device and the OBFCM Data Service for the main functional objective of the system, which is the periodic transfer of the OBFCM values, from the vehicle to the EC.

Figure 13: OTA Device becomes operational



Source JRC 2022

OTA Device sends MQTT message to OBFCM Data Service

While the OTA Device is at the “initialized'' state, it periodically connects to the MQTT broker of the OBFCM Data Service and retrieves the latest configuration settings. After the OTA Device transitions to an operational state, it stays disconnected from the MQTT broker most of the time and only connects when it is time to exchange data. The OTA Device periodically checks if the time has come to check for a configuration update or to send OBFCM data by looking at its locally stored configuration.

OBFCM Data Service sends MQTT message to OTA Device

Once a vehicle / OTA Device has been cleared to report OBFCM data (i.e., is moved from the “deny-list” to the “allow-list”), a retained configuration status message will be published by an OBFCM Data Service component to the respective MQTT broker. This message will contain all necessary information such as data acquisition frequency, data transfer frequency, etc., as well as a command for the OTA Device to enter “operational” state. The message will be received as soon as the OTA Device connects to the MQTT broker.

OTA Device updates its state and acquires data from OBFCM device

When the OTA Device receives this MQTT message it goes into “operational” state. This means that it periodically reads data from the vehicle’s OBFCM device and saves it to local storage.

OTA Device sends data to OBFCM Data Service

The OTA Device periodically reads the pre-stored OBFCM data from local storage and publishes it via MQTT to the OBFCM Data server.

OBFCM Data Service stores data and OTA Device refreshes storage

The MQTT broker will send a protocol-level publish acknowledgement (PUBACK) to the OTA Device that the OBFCM data transfer message was received. The OBFCM Data server will subsequently store the incoming data in a time-series database.

Note that the PUBACK message is not a distinct MQTT confirmation message published by the OBFCM Data Service, as in the case of the registration confirmation. It is a low-level acknowledgment message sent by the MQTT broker. This acknowledges that the previous data publish message has reached the MQTT broker but it does not mean that the data have been successfully stored. When the OTA Device receives the ACK it will delete the respective data from local storage.

To address connectivity failures, the OBFCM data transfer is re-attempted until a PUBACK message is received from the MQTT broker indicating that the data have reached the other side.

4.3 AWS infrastructure deployment

This Proof-of-Concept implementation utilized AWS IoT infrastructure services [13] to enable secure and scalable MQTT communication between OTA Devices and off-board server systems. There are two MQTT brokers, one used by the OTA Registration Service that handles the device registration process and one used by the OBFCM Data Service that handles the OBFCM data transfers and device reconfigurations.

To prepare PoC components to work with the AWS IoT infrastructure, a set of tasks regarding identity management and access management must be carried out. The first involves issuing X.509 certificates and the second involves Just-In-Time-Provisioning templates and MQTT access policy documents.

4.3.1 Identity management for OTA Registration Service and OBFCM Data Service

Our OTA Registration Service and OBFCM Data Service applications act as MQTT clients when connecting to the MQTT broker. The TCP connection between the MQTT broker and MQTT clients needs to be secure. This requires mutual authentication - both sides of the connection will verify each other's identity. MQTT clients need to present digital certificates which the MQTT broker and vice versa trust - the broker’s certificate must be trusted by the client.

So, when the OTA Registration Service application connects to the AWS IoT Core service it needs to present an X.509 certificate which is signed by a Certificate Authority trusted by AWS IoT Core. The AWS IoT service allows cloud infrastructure administrators to manually issue X.509 certificates which can be used for this purpose. We

used this facility to issue two such certificates (one per application) and placed them in the local resource folder of each application. This allows the certificate to be loaded at runtime and used by the application's MQTT client library.

4.3.2 Access management for OTA Registration Service and OBFCM Data Service

The MQTT policy document for the OTA Registration Service specifies that the server may:

- publish messages on the "registerServer'' topics of all OTA Devices, and
- subscribe to and receive messages from the "registerDevice" topics of all OTA Devices.

The "registerDevice" topic is where the OTA Device sends registration requests and the "registerServer" topic is where the OTA Registration Service posts retained messages reflecting the device registration status.

Similarly, the MQTT policy for the OBFCM Data server specifies that the server may:

- publish to the "configServer" topics of all OTA Devices, and
- subscribe to and receive messages from all devices' "dataDevice" topics.

The "configServer'' topic is where the OBFCM Data server sends the configuration data. The "dataDevice'' topic is where the OTA Device sends OBFCM data.

4.3.3 Identity management for OTA Devices

The exchange of MQTT messages between the OTA Devices and the MQTT broker offered by the AWS IoT Core service also requires mutual identity authentication with X.509 certificates.

As mentioned earlier in the report, we developed an application to act as a CA so we can issue X.509 certificates for large batches of OTA Devices. For this PoC implementation, the CA application has a self-signed certificate. This certificate has been added to the AWS IoT list of trusted CAs. As a result, all of the X.509 OTA Device certificates which are issued by this specific CA application are accepted by the MQTT broker on AWS as trusted proof of identity. The certificate of each device is placed in the local resource folder of the Fleet Simulator application.

4.3.4 Access management for OTA Devices

For access management on the MQTT broker side, we employed the mechanism of Just-In-Time-Provisioning (JITP) [14] offered by AWS IoT. JITP allows IoT devices to be provisioned with access rights for using AWS IoT Core on the fly when they first attempt to connect to the MQTT broker. In other words, devices can connect to the MQTT broker without having their X.509 certificate already registered with AWS IoT, which is the case for our server applications (OTA Registration Service and OBFCM Data Service).

To utilize JITP the following are needed: a) the X.509 certificate of the Certificate Authority which signs the OTA Devices' certificates must be registered with AWS IoT, and b) a provisioning template defining the MQTT topic access policy must also be submitted to AWS IoT. The provisioning template which is registered with AWS IoT specifies access rights on MQTT topics for all OTA Devices connecting to the MQTT broker.

The runtime process of JITP follows these steps: whenever an OTA Device connects to the MQTT broker for the first time, AWS IoT checks whether a listed (trusted) CA has signed the device's certificate. If so, AWS IoT registers the device's certificate, creates an IoT thing, creates a policy document based on the provisioning template of the CA. It then attaches the policy to the device certificate and the certificate to the IoT thing. Lastly, the certificate is activated and the device can continue using the MQTT broker. If any of these steps fail because of misconfiguration / bad credentials the device is disconnected from the MQTT broker.

Using our CA application, we issued certificates for our OTA Devices and we placed the OTA Device ID in the certificate's subject Common Name field. This allowed us to specify, in the provisioning template, the dedicated MQTT topics which an OTA Device can access by indirectly including its ID in the topic's name since the AWS IoT Core service can read the device ID from the device's certificate.

We set up the OTA Registration Service and the OBFCM Data Service's MQTT infrastructure in two AWS regions (Europe - Frankfurt / eu-central-1 and Europe - Ireland / eu-west-1). Since we are employing two different MQTT brokers the JITP setup process needed to be done twice, once for each MQTT broker. We registered the X.509 certificate of the Certificate Authority on both AWS regions and provided two provisioning templates, one for accessing the OTA Registration Service's broker and one for accessing the OBFCM Data server's broker.

5 Scaling up to collect data from over 100 million vehicles

5.1 Scalable cloud service architecture

The first iteration of the Proof-of-Concept implementation described in Chapter 4 was designed to demonstrate feasibility rather than scalability.

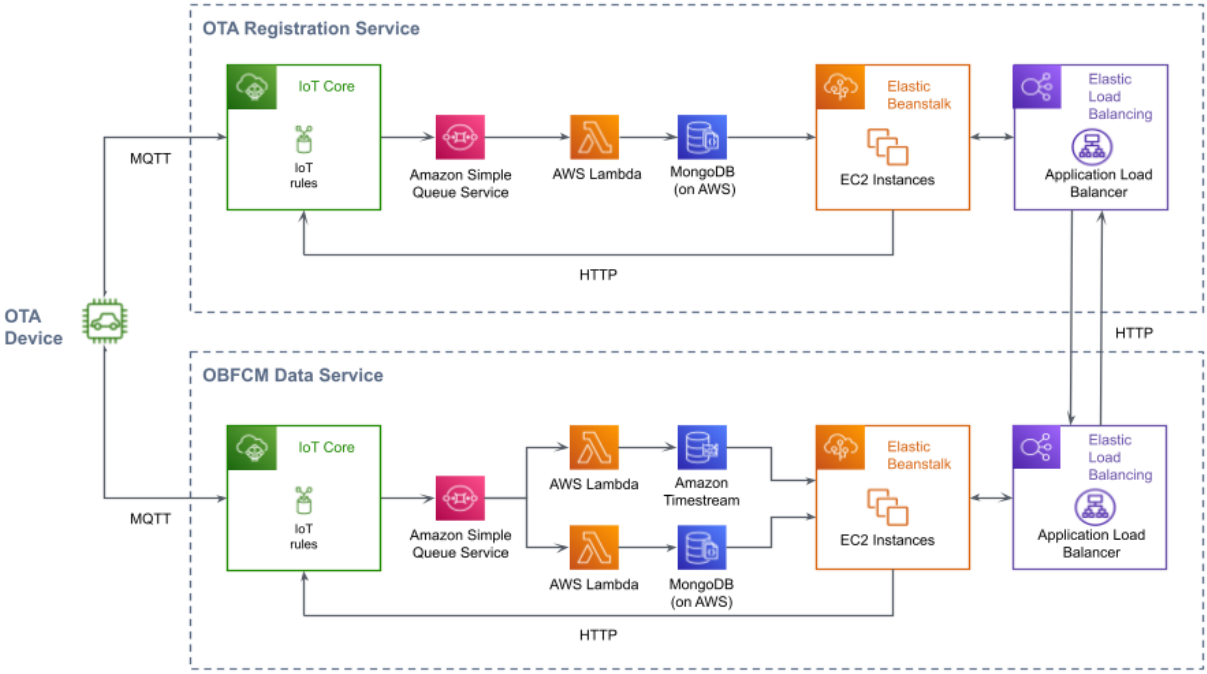
In the second iteration of the PoC we expanded the architecture and evolved the implementation to allow operation at a large scale. We developed a cloud service architecture on top of AWS to support large-scale workloads and off-loaded some of the processing done inside our applications to AWS services. Most importantly, we introduced highly-scalable message queueing and asynchronous lambda functions to lift the heavy load of concurrent connections to the MQTT broker that will be arriving from tens of millions of vehicles. The OTA Registration and OBFCM Data Service applications were also redesigned for horizontal scaling, to accommodate high volumes of device-to-cloud communication and intra-server communication.

There are several different ways to design a scalable architecture on top of AWS cloud services. Below is the list of services that we used in our prototype:

- AWS IoT Core - Scalable MQTT broker service
- AWS SQS - Scalable message queuing system
- AWS Lambda - Scalable event-driven execution of data processing functions
- AWS Timestream - Scalable time-series database
- AWS Beanstalk - Service for deploying and scaling web applications
- MongoDB deployed on AWS - Document database hosted on a dedicated cluster

A model of the interaction between those AWS services is illustrated in Figure 14 below.

Figure 14: Cloud architecture supporting large scale fleet-wide data collection



Source JRC 2022

As shown in the architecture diagram of Figure 14, the order of processing is as follows:

1. Incoming MQTT connections from vehicles are routed to the AWS IoT Core service. This is the case for both the OTA Registration Service and the OBFCM Data Service.
2. The content of the MQTT messages arriving at the **AWS IoT Core** service is evaluated against preset **IoT Rules**[4]. This results in the messages being copied onto a scalable message queue implemented with the **AWS Simple Queue Service**[5].
3. **AWS Lambda functions**[6] are polling the SQS queue for incoming messages. This triggers the execution of simple and automatically scalable functions which transfer the data onto the appropriate data storage.
4. In the case of the OTA Registration Service the storage of static data about the vehicle and its OTA Device is done in a **MongoDB Atlas**[7] document database. The same happens with static OTA Device data arriving at the OBFCM Data Service. The dynamic OBFCM value data are stored in a specialized time-series database - **Amazon Timestream**[8].
5. The data is read from MongoDB and Timestream storage and is processed by the two different kinds of server applications we have already discussed extensively in this report - the registration server and the data server. Each application runs on several virtual servers in parallel (multiple **Elastic Compute Cloud** instances[9]). If necessary, new EC2 instances can be created automatically to serve additional demand.
6. Intra-server communication between the registration server and the data server is performed via HTTP REST calls. An **Application Load Balancer** helps distribute the weight of the workload evenly between the EC2 instances of each server which run in parallel.
7. The registration status messages sent by the OTA Registration Service to the OTA Device and the configuration status messages sent by the OBFCM Data Service to the OTA Device are delivered to the devices to IoT Core as retained messages via HTTP REST calls.

5.2 Data communication footprint per vehicle

The total volume of data to be exchanged per vehicle, per year, is determined by the frequency of communication events in the span of a year and the size of the data exchanged per communication event.

In the analysis that follows we examine the following scenario: Once per month the OTA Device acquires a snapshot of values from the OBFCM device, and once per quarter-year[10] it transfers the acquired snapshots in a single batch. When connecting to the OBFCM Data Service to transfer the data the OTA Device also receives the latest configuration status. Table 13 below presents the frequency of events.

Note that OTA Device registration requests are not repeated on an annual basis. This type of device-to-cloud communication event happens only once in the lifetime of the OTA Device: when a device in the form of a hardware and/or software is installed on a vehicle. In the sake of simplicity, we do not consider communication events triggered by the replacement of the OTA Device on a vehicle with a new component. This could be factored into the analysis at a vehicle fleet level, rather than at an individual vehicle level.

[4] https://docs.aws.amazon.com/iot/latest/developerguide/iot-rules.html
[5] https://docs.aws.amazon.com/sqs/index.html
[6] https://docs.aws.amazon.com/lambda/latest/dg/with-sqs.html
[7] https://www.mongodb.com/
[8] https://aws.amazon.com/timestream/
[9] https://aws.amazon.com/elasticbeanstalk/
[10] This assumption applies through the lifetime of the vehicle for simplicity. However a dynamically adjustable approach were newer vehicles communicate more frequently the OBCM values compared to older ones would probably make more sense in terms of monitoring, and robustness of the consumption values recorded.

Table 13. Frequency of communication events

Type of communication event	**Description**	**Frequency of events / year**
Request for new OTA Device registration	OTA Device connects to the OTA Registration service and requests to enrol into the data collection network	Not applicable (only once per OTA Device installation)
Acquisition of OBFCM data (on-board operation)	OTA Device reads values from OBFCM device and stores them locally	12
Transfer of data	OTA Device connects to the cloud and transfers the previously stored OBFCM values (lifetime fuel and lifetime distance only)	4
Retrieval of configuration status	OTA Device connects to the cloud-based OBFCM Data Service and receives latest configuration status/instructions	4

Source: JRC, 2022

Table 14 presents the data payload size being exchanged per communication event when this data is in unencrypted form. To calculate the data payload size for the data transfer event, we assume that each data transfer (once per quarter) includes three OBFCM value snapshots (one taken per month).

Table 14. Data payload size

Message type	**Volume per message - unencrypted (Bytes)**	**Volume per message - unencrypted (Kilobytes)**	**Annual frequency**	**Total volume per message - unencrypted (Bytes)**	**Total volume per message - unencrypted (Kilobytes)**
Request for new OTA Device registration	438 B	0.43 KB	-	-	-
Transfer of data	377 B	0.37 KB	4	1508 B	1.48 KB
Retrieval of configuration status	1178 B	1.15 KB	4	4712 B	4.60 KB
	1993 B	**1.95 KB**	**8**	**6220 B**	**6.07 KB**

Source: JRC, 2022

As shown in Table 14 above, **the total unencrypted data payload is approximately 6 KB per vehicle per year**. However, the data must be secured while in transit. The use of mechanisms to secure the transport layer such as data encryption and certificate-based authentication adds to the overall data volume of the communication.

Table 15. Encrypted data size

Total encrypted data volume, per vehicle, per year	**Bytes**	**Kilobytes**
TLS cipher suite used: TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	**82,671 B**	**80.73 KB**

Source: JRC, 2022

Using the Wireshark network protocol analyser[11], we measured the total volume of the fully encrypted and authenticated data communication during a (simulated) full year of vehicle-to-cloud communications. The result as shown in Table 15 is **approx. 80 KB per vehicle per year**, given the assumptions, stated earlier regarding frequency of communication events per year.

The cipher suite that was used during the measurement with the network protocol analyser was TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256[12]. The use of a different Transport Layer Security cipher suite would mean that a different set of algorithms would be used for key exchange, encryption and hashing. In turn, this would result in a different data footprint.

Based on what was considered above it appears that the footprint of the data communication, per vehicle, per year, is negligible. 80 KB is approximately the size of an email message without attachments.

Note that we only considered the size of the lifetime fuel and distance values when calculating the size of the data transfer message. If we were to include the six additional OBFCM lifetime values which are relevant for hybrid electric vehicles (see section 1.1) the unencrypted data payload of 6KB would likely increase. Expanding the message structure could lead to further increases, however, **the total fully encrypted data footprint per vehicle would likely not exceed 1MB per year**.

One additional dimension that is relevant to the data communication footprint of the OTA Device is the duration of vehicle-to-cloud connectivity. Irrespective of how much data is transferred through an internet connection, it is also important to know how long a network connection between the vehicle and the cloud needs to be kept open. According to the off-board communication protocol presented in chapter 3 the OTA Device does not stay connected to the cloud permanently. It only connects when it is time for a periodic data transfer operation or periodic configuration check and disconnects immediately afterwards. This is expected to be a few seconds per communication event. We should also account for the potential overhead to be added by OTA Device reconnection attempts, which is expected to happen if server connection limits are reached. OTA Devices should be expected to overcome connectivity problems by applying an exponential back-off and retry algorithm. In either case, the total duration **of vehicle-to-cloud connectivity per year, at the individual vehicle level, is likely to be no more than 1 minute**.

5.3 Vehicle fleet size assumptions

To estimate the annual cost of operating the cloud service infrastructure outlined earlier, we first need to estimate the number of communication events expected to occur at the level of the entire vehicle fleet. This figure is determined by the number of OTA-equipped light-duty vehicles (LDVs) participating in OBFCM data collection.

As a starting point, we made some assumptions regarding the size of the OTA-equipped vehicle fleet and how it will evolve over 10 years from the introduction of OTA data collection into regulation.

Simplifying assumptions:

- Constant number of 14.8 million new LDVs registered in the EU per year, throughout the 10-year period. This figure is based on the number of passenger vehicles and light commercial vehicles which were registered in the EU in 2019[13].
- The percentage of new LDV owners who consent to participate in over the air collection of OBFCM data is a constant 70% throughout the 10 years. This assumption is not based on any evidence, it is speculative.
- The percentage of the LDV fleet that succeeds in communicating with the EC cloud services at least once per year is 98%. This is equal to the 2019 rate of 4G coverage in rural areas of the EU[14]. This figure remains constant throughout the 10-year period.

Based on those assumptions, the LDV fleet size figures relevant for cost estimation are shown in Table 16.

[11] https://www.wireshark.org
[12] Keys exchange performed with ephemeral Elliptic Curve Diffie Hellman (ECDHE), authentication with RSA, bulk encryption with AES Galois Counter Mode with 128-bit key size (AES_128_GCM) and hashing with SHA-256.
[13] https://www.acea.auto/figure/2019-motor-vehicle-registrations-in-europe-by-country
[14] https://digital-strategy.ec.europa.eu/en/library/connectivity-european-gigabit-society-brochure

Table 16. Assumptions on vehicle fleet size in 10-year period

LDV fleet size metrics relevant for cost estimation	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
New OTA-equipped LDvs registered for circulation in EU	14.8M	14.8M	14.8M	14.8M	14.8M	14.8M	14.8M	14.8M	14.8M	14.8M
Cumulative size of OTA-equipped LDV fleet in circulation	14.8M	29.5M	44.3M	59.0M	73.8M	88.5M	103.3M	118.0M	132.8M	147.5M
Percentage of new vehicle owners who consent to data collection	70%	70%	70%	70%	70%	70%	70%	70%	70%	70%
Newly registered LDvs whose owners consent to data collection	10.3M	10.3M	10.3M	10.3M	10.3M	10.3M	10.3M	10.3M	10.3M	10.3M
Cumulative number of LDvs whose owners consent to data collection	10.3M	20.7M	31.0M	41.3M	51.6M	62.0M	72.3M	82.6M	92.9M	103.3M
Percentage of fleet that successfully connects to EC cloud services at least once per year	98%	98%	98%	98%	98%	98%	98%	98%	98%	98%
Newly registered LDvs that connect to transfer data for the first time	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M
Cumulative size of LDV fleet transferring data at least once per year	10.1M	20.2M	30.4M	40.5M	50.6M	60.7M	70.8M	81.0M	91.1M	101.2M

Source: JRC, 2022

5.4 Communication frequency assumptions

On the basis of this fleet size estimation, we can calculate the frequency of Over-the-Air communication events to take place per year, assuming, once more, that there is one data transfer event taking place per quarter (four times per year). As previously mentioned, a dynamic sampling approach emphasizing the first years of vehicle operation with more frequent sampling and less frequent sampling later in vehicles' life would be a reasonable approach not investigated in the report for simplicity. Such an approach would not significantly affect the total number of communication events over the decade. We further assume that the vehicle's OTA Device is checking for new configuration updates with the same frequency.

As shown in Table 17 below we assume approximately 10 million of OTA Device registration requests per year, from owners of new vehicles who consent to data collection. In the interest of simplicity this figure is assumed to remain constant throughout the 10-year period. **By Year 10 we assume approximately 400 million data transfers (and configuration updates) taking place annually from a vehicle fleet of over 100 million vehicles.**

Note that we have not made detailed estimates of some secondary triggers of communication events at this stage of the analysis, which are not primary contributors to the workload and are therefore not among the primary cost drivers.

Table 17. Assumptions on communication frequency in 10-year period

Fleet-wide vehicle-to-cloud communication events	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Device registration requests	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M	10.1M
Data transfers	40.5M	81.0M	121.4M	161.9M	202.4M	242.9M	283.4M	323.8M	364.3M	404.8M
Configuration checks	40.5M	81.0M	121.4M	161.9M	202.4M	242.9M	283.4M	323.8M	364.3M	404.8M

Source: JRC, 2022

Secondary vehicle-to-cloud communication events

Data transfers which are the result of a **manually-induced OTA Communication Test during the Periodic Technical Inspection** of a vehicle (see section 1.3.3). In most EU member states, vehicles purchased in Year 1 will be undergoing PTI in Year 4. This could potentially increase the number of data transfer communication events happening from Y4 onwards by 25% (5 data transfer events per vehicle per year instead of 4 events).

Secondary inter-cloud communication events

The OTA Registration Service will be receiving messages from different types of vehicle registrar organizations (manufacturers or transport authorities) to update its vehicle registry information (see section 4.2.2). For instance, some legal entity that is tasked with **managing the vehicle owner's consent** (this is outside the scope of the OTA data collection network) will need to notify the OTA Registration Service when a vehicle owner has withdrawn their consent for data collection or has provided new consent.

Similarly, the OTA Registration Service will need to be notified when a vehicle is registered in a different EU member state. It is not clear at this stage if the same should happen for every transfer of vehicle ownership, even if that takes place in the same country. The **average passenger car has 3 to 4 owners during its lifetime** [15]. A communication overhead should be expected as a result.

Secondary intra-cloud communication events

The progressively increasing vehicle-to-cloud and inter-cloud communication events will also drive an increase in intra-cloud communication events. The latter are messages exchanged between the OTA Registration Service and the OBFCM Data Service following the protocol explained in sections 4.2.1 and 4.2.2.

5.5 Estimation of operational cost for cloud infrastructure

Moving forward, we created a cost model incorporating each of the AWS services introduced in section 5.1 by recreating the formulas provided by the AWS Pricing Calculator[15]. Where needed, we made further assumptions regarding the formula inputs based on our experience from developing the Proof of Concept.

AWS has the concept of a Region, a physical location where the company clusters several data centres. This is relevant for cost calculation because the pricing for the same AWS service may vary across Regions. In calculating the costs here, we used the publicly available price list of AWS for services provided through its EU Frankfurt data centre.

Table 18 and Table 19 provide estimates of the operating costs for the OTA Registration Service and OBFCM Data Service on Year 1, Year 5 and Year 10. The figures quoted are based on the publicly available price list of AWS and have been converted to Euro at the present-day currency exchange rate of 0.86 Euro per US Dollar.

[15] https://calculator.aws/#/

Table 18. Estimated annual cost of AWS services for OTA Registration Service

Estimated annual operating cost of OTA Registration Service	**Y1**	**Y5**	**Y10**
AWS IoT Core	€60	€60	€60
AWS SQS	€7	€7	€7
AWS Lambda	€105	€105	€105
AWS Timestream	-	-	-
AWS Beanstalk	€553	€553	€553
MongoDB Atlas	€1,507	€4,068	€7,835
Total	**€2,231**	**€4,793**	**€8,560**

Source: JRC, 2022

As per our fleet size assumptions, the OTA Registration Service will be enrolling a constant 10.1 million new vehicles/OTA Devices per year, throughout the ten-year period of interest. The only factor that increases the operational cost from Y1 through to Y10 is the gradually accumulating volume of static vehicle-related data which is placed on storage - and the associated cost of the managed cloud-based database service.

Table 19. Estimated annual cost of AWS services for OBFCM Data Service

Estimated annual operating cost of OBFCM Data Service	**Y1**	**Y5**	**Y10**
AWS IoT Core	€297	€1,483	€2,966
AWS SQS	€28	€139	€279
AWS Lambda	€627	€3,413	€6,894
AWS Timestream	€25	€379	€1,388
AWS Beanstalk	€900	€1,596	€1,943
MongoDB Atlas	€1,507	€4,068	€7,835
Total	**€3,384**	**€11,077**	**€21,305**

Source: JRC, 2022

The operating cost of the OBFCM Data Service follows a different pattern. The cost figures shown in Table 19 above are based on the assumption that by Y10 there will be over 100 million LDVs which are transferring data Over-the-Air, growing tenfold from Y1. At the same time the operating cost between Y1 and Y10 increases by a factor of 5.6.

The total cost of operating the OBFCM Data Service to collect data from over 100 million vehicles by Year 10 could be a rather modest €21,305 per year. With the addition of €8,560 for the OTA Registration Service the total annual operating cost is under €30,000.

Note that this estimate is produced without taking into consideration the AWS cloud infrastructure costs for:

— Replication of the production infrastructure environment to address the needs of software lifecycle management, i.e., separate development and staging environments

— Infrastructure redundancy, i.e., identical production-grade systems deployed in different Availability Zones in the same AWS Region to guarantee uptime.

— Data backup services, e.g., cold storage.

Considering those additional requirements, and not excluding the possibility that more requirements surface through feedback from the relevant EC departments and agencies, it could be possible that the estimated figure increases significantly. However, given the assumptions stated earlier, it appears unlikely that the total AWS operating costs for OTA data collection could exceed €100,000 per year.

This does not cover development expenses but rather the operational costs of such a system. From a conservative perspective, an increase of €100,000 could be assumed to account for possible hidden expenses.

6 Conclusions and further research

The European Commission has the mandate to collect and process the data recorded by On-board Fuel and/or Energy Consumption Monitoring Devices for monitoring real-world CO_2 emissions of road vehicles.

Regulation (EU) 2019/631 specified three different routes through which the Commission shall regularly be collecting the OBFCM measurement data: "*from manufacturers, national authorities or through direct data transfer from vehicles*". Commission Implementing Regulation (EU) 2021/392 introduced a legal requirement for manufacturers and Member States to start collecting data and report it to the Commission and the EEA on an annual basis. Data cannot be collected for both manufacturers and Member State authorities if the vehicle owner refuses to make it available.

This report investigated the technical requirements and solutions for the third route, i.e., direct data transfer from vehicles to the EC, without intermediaries. **Compared to data collection through manufacturers or the Member States, the collection of fleet-wide data through direct data transfer from vehicles has some distinct relative advantages:**

- **Faster lead-time for data collection:** Direct data collection can be activated as soon as a new vehicle leaves the production floor and it can start working as soon as the vehicle is registered in an EU member state and the owner agrees. In contrast, collection via manufacturers' authorized repairers will require up to 12 months for the data to be collected for the first time and another few months until it is forwarded to the Commission. While for most EU member states, collection via PTI agencies can start no earlier than 4 years from a vehicle's registration.
- **Data collection as frequently as needed:** Direct data transfer allows for frequent data sampling. Higher data resolution will enable researchers and stakeholders to understand the contributing factors to variability in CO_2 emissions in ways that had not been possible before. At the same time, in direct transfer, the frequency of acquiring data from the OBFCM device and transferring the data are independent parameters which can be dynamically configured at zero cost. This is not the case with data collection from PTI agencies or from manufacturers' authorized partners where data snapshots will be collected annually. Also, in the case of data collection via manufacturers' authorized repairers the supply of data is expected to gradually become sparser for vehicles whose warranty period has ended and whose owners turn to independent repairers, making CO_2 monitoring more reliant on PTI-based data collection.
- **Assurances for data reliability:** Increasing the number of entities that are involved in the processing of data along a communication path makes data errors more likely to appear and the causes of errors more difficult to trace back to their root. OTA data transfer directly to the EC reduces the risk of system or human error because it minimizes the number of intermediate data processing/forwarding systems and eliminates manual intervention. It also reduces the surface space for malicious intervention by individuals or organizations aimed at tampering the data.
- **Assurances for data privacy**: Decreasing the number of intermediaries means that there will be fewer entities to whom personal data will be exposed. Also, direct transfer of data makes it possible to universally and uniformly enforce personal data protection measures and to introduce pseudonymization for Vehicle Identity Numbers as an additional measure of data privacy protection.

Given the above, direct data transfer is considered a method of OBFCM data collection that is complementary to the PTI-based and manufacturer-based routes in important dimensions. In this context, a technical solution framework was developed which conceptualizes a method of direct data transfer that is secure, privacy-preserving, tamper-resistant, scalable and future-proof. The solution framework introduces the concept of an OTA Device, i.e., an on-board software or hardware unit that is simultaneously integrated with both on-board and off-board systems. On-board integration concerns how the OTA Device communicates via the vehicle network with the onboard OBFCM device and other onboard ECUs. Off-board integration concerns how the OTA Device communicates via the mobile network to remote cloud services under the European Commission's responsibility.

The report focused on off-board communication exclusively. We introduced an off-board communication protocol for OTA transfer of OBFCM data which is defined based on MQTT - a standard with growing adoption by the automotive industry (ISO/IEC 20922). The communication protocol that we defined enables the integration of the three main technology components of the OTA data collection network: a) the OTA Devices, b) the OTA Registration Service, and c) the OBFCM Data Service. We discussed functionality, the context of use, frequency/volume of use, and operational supervision for each of those components.

In our definition of the MQTT-based communication, we covered the MQTT topic schema and respective topic access rights, the MQTT message types and the message structure. Following, we presented a proof-of-concept implementation of this protocol, which was developed to demonstrate the feasibility of the solution. We demonstrated how such a system can be built based on modern web technologies combined with secure and scalable IoT infrastructure services provided by a widely used cloud services provider, AWS.

The report detailed the five different applications that make up the PoC components, i.e., OTA Registration Service, OBFCM Data Service, Fleet simulator, Certificate Authority simulator, and Workflow demonstrator. We discussed functionality, technology stack and implementation decisions. It also described the three most important processes carried out in the OTA data collection network and implemented in our PoC, walking through these processes step by step, with the help of UML sequence diagrams. Lastly, we elaborated on how identity management and access management were realized on AWS infrastructure services.

The PoC implementation demonstrates that the approach we have put forward is technically feasible. We have developed a fully working system that validates the solution approach from a functional viewpoint. The main characteristics of the solution are:

- **Data reliability and privacy by design**, made possible through public-key cryptography (digital certificates, encryption, cryptographic signatures, PKI system) and tamper protection measures in OTA Devices (Trusted Platform Module).
- **Anonymous data collection**, achieved through pseudonymization of the data source (associating the collected OBFCM data with an OTA Device ID instead of the VIN) and data segregation (keeping vehicle data physically separate from vehicle identifiers).
- **Scalability to over 100 million vehicles**, by extending the cloud service architecture with AWS infrastructure services (asynchronous message queueing, serverless Lambda functions, horizontal application scaling for concurrent processing).
- **Low communication footprint**, estimating the footprint of the OTA data communication to be less than 1MB per vehicle per year.
- **Manageable operational costs**, estimating the total cloud infrastructure costs for OTA data collection from 100 million vehicles to not exceed €100,000 per year.

As future work, the scaled-up PoC implementation presented in chapter 5 should be tested to improve its design by running a large-scale simulation with millions of vehicles. Running this type of simulation will allow us to better understand the architecture's potential limitations and validate the cost estimation model.

In the scope of a large-scale simulation, we will also examine extensions to our MQTT topic schema to allow for configuration updates to be broadcast across large predefined segments of vehicles (in addition to the point-to-point communication pattern we currently employ in configuration updates).

Moreover, a future activity could focus further on applications related to other emissions such as On-board Monitoring (OBM), spanning not only CO_2 but also other pollutants. Another critical question is the alignment of this solution approach with other technical solutions being studied today related to OBFCM data. The same data collection architecture described in this report is directly applicable for collecting and analysing large-scale data from vehicles.

References

[1] Commission Regulation (EU) 2018/1832 of 5 November 2018 amending Directive 2007/46/EC of the European Parliament and of the Council, Commission Regulation (EC) No 692/2008 and Commission Regulation (EU) 2017/1151 for the purpose of improving the emission type approval tests and procedures for light passenger and commercial vehicles, including those for in-service conformity and real-driving emissions and introducing devices for monitoring the consumption of fuel and electric energy. Available at: https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX:32018R1832

[2] Regulation (EU) 2019/631 of the European Parliament and of the Council of 17 April 2019 setting CO_2 emission performance standards for new passenger cars and for new light commercial vehicles, and repealing Regulations (EC) No 443/2009 and (EU) No 510/2011. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019R0631

[3] Regulation (EU) 2019/1242 of the European Parliament and of the Council of 20 June 2019 setting CO_2 emission performance standards for new heavy-duty vehicles and amending Regulations (EC) No 595/2009 and (EU) 2018/956 of the European Parliament and of the Council and Council Directive 96/53/EC. Available: https://eur-lex.europa.eu/eli/reg/2019/1242/oj

[4] Kourtesis, D., and Fontaras, G. “Technical requirements for OTA transfer of OBFCM data - JRC proposal on Over-the-Air data transfer”. Presented at the 6th meeting of the European Commission OBM-Task Force, 15 December 2020.

[5] FIA Report. Creating a level playing field for vehicle data access: Secure On-board Telematics Platform Approach. February 2021.

[6] ISO 20078-1:2019 Road vehicles — Extended vehicle (ExVe) web services. February 2019.

[7] ISO/IEC 20922 Message Queuing Telemetry Transport - MQTT.

[8] ACEA Report - Economic and Market Report - EU Automotive Industry Full-year 2019. May 2020. Available: https://www.acea.auto/files/Economic_and_Market_Report_full-year_2019.pdf

[9] ACEA Report - Vehicles in Use, Europe 2021, January 2021 https://www.acea.auto/publication/report-vehicles-in-use-europe-january-2021/

[10] MQTT Version 3.1.1. December 2015. Available: http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/mqtt-v3.1.1.html

[11] HiveMQ. Introducing the MQTT Protocol - MQTT Essentials. Available: https://www.hivemq.com/blog/mqtt-essentials-part-1-introducing-mqtt/

[12] Amazon Web Services. Designing MQTT Topics for AWS IoT Core. May 2019.

[13] Amazon Web Services. AWS IoT - IoT services for industrial, consumer, and commercial solutions. Available: https://aws.amazon.com/iot/

[14] Amazon Web Services. Just-in-time provisioning. Available: https://docs.aws.amazon.com/iot/latest/developerguide/jit-provisioning.html

[15] Vanherle, K., and R. Vergeer. "Data gathering and analysis to improve the understanding of 2nd hand car and LDV markets and implications for the cost effectiveness and social equity of LDV CO_2 regulations-Final Report for: DG Climate Action." TML/DG Climate Action (2016).

List of abbreviations and definitions

ACEA	European Automobile Manufacturers' Association
ACK	Acknowledgement
AES	Advanced Encryption Standard
AIRC	Association Internationale des Réparateurs en Carrosserie
API	Application Programming Interface
AWS	Amazon Web Services
CA	Certificate Authority
CAN	Controller Area Network
CECRA	European Council for Motor Trades and Repairs
CLI	Command Line Interface
CO2	Carbon Dioxide
CSR	Certificate Signing Request
DB	Database
EC	European Commission
EC2	Amazon Elastic Compute Cloud
ECDHE	Elliptic Curve Diffie-Hellman Ephemeral
ECU/TCU	Engine Control Unit/Telematics Control Unit
EEA	European Environmental Agency
EU	European Union
FIA	Federation Internationale De L'Automobile
FIGIEFA	Fédération Internationale des Grossistes, Importateurs & Exportateurs en Fournitures Automobiles
GCM	Galois Counter Mode
HMI	Human-Machine Interface
HTTP	HyperText Transfer Protocol
ID	Identity
IEC	International Electrotechnical Commission
IP	Internet Protocol
ISC	In Service Conformity
ISO	International Organization for Standardization
JITP	Just In Time Provisioning
JRC	Joint Research Centre
JSON	JavaScript Object Notation
LDV	Light Duty Vehicles
MQTT	Publish/subscribe messaging transport protocol standardized by ISO/IEC 20922:2016 (the acronym was formerly standing for Message Queuing Telemetry Transport)
OBD	On-board Diagnostics
OBFCM	On-board Fuel and/or energy Consumption Monitoring
OBM	On-board Monitoring
OEM	Original Equipment Manufacturer
OTA	Over The Air
PKI	Public Key Infrastructure
PoC	Proof of Concept
PTI	Periodic Technical Inspection
REST	REpresentational State Transfer
RSA	Rivest–Shamir–Adleman public key encryption cryptosystem
SHA256	Secure Hash Algorithm novel hash function computed with eight 32-bit words
SQS	Simple Queue Service
TCP	Transmission Control Protocol
TLS	Transport Layer Security
TPM	Transactions Per Minute
UML	Unified Modeling Language
VIN	Vehicle Identification Number
WLTP	Worldwide harmonized Light vehicle Test Protocol

List of boxes

List of figures

List of tables

Annexes

Annex 1. OTA Device functions

The table below indicates the source and destination of the information flow taking place when each device function is exercised, in both on-board and off-board communication.

Interface	Function	Information source	Information destination	Communication initiator	Communication objective
Off-board communication	Device Registration	OTA Device	OTA Registration Service	OTA Device	Send registration request
	Device Initialization	OBFCM Data Service	OTA Device	OTA Device	Receive configuration (initial)
	Device Configuration	OBFCM Data Service	OTA Device	OTA Device	Receive configuration
	Data Transfer	OTA Device	OBFCM Data Service	OTA Device	Send OBFCM data
Both On-board and Off-board (depends on implementation)	Device Provisioning	Data storage and/or external systems	OTA Device	OTA Device	Receive security-related assets
On-board communication	Data Acquisition	OBFCM device	OTA Device	OTA Device	Receive OBFCM measurements
	Device Status Push	OTA Device	HMI or other ECUs	OTA Device	Send status info
	Device Status Pull	OTA Device	OBD scan tool, HMI or other ECUs	OBD scan tool, HMI or other ECUs	Receive status information
	OTA Communication Test	OTA Device	OBFCM Data Service	OBD scan tool	Send OBFCM data (manually induced)

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-30964-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/46158

ISBN 978-92-76-46614-7