



WAVESTONE

GRIMALDI STUDIO
LEGALE



TNO



KOMIS

3rd Workshop
Pilot on **Fair and equal data sharing for cooperative, connected and automated mobility**
Presentation of planned updates to the architecture, additional features and services

Big Data and B2B platforms: the next big opportunity for Europe
EASME/COSME/2018/004

Konstantinos Tzamaloukas
Vaggelis Karakolis
ICCS-NTUA

EASME - European Commission
Executive Agency for Small and
Medium-sized Enterprises



Online Workshop
23 June, 2020

Two major new enhancements of the architecture

Bilateral Communication between vehicles and OEMs

Elaborating Anonymization

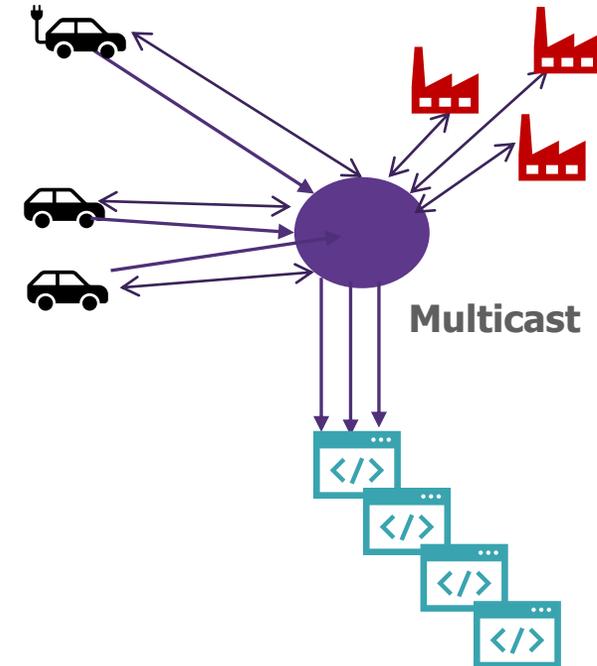


Bilateral Communication between vehicles and OEMs – The Multicast Data Flow Model

- The Multicast Data flow model was **preferred by most of the WS2 participants**
- The upcoming release of Shared Server Architecture will support the Multicast data flow Model

Specifically:

- The Shared Server receives a continuous flow of data from the vehicles in order to promote them to service providers
- OEMs Server will receive a continuous flow of data from the vehicles
- OEM can push updates and notifications to the vehicles
- OEMs will receive data from the vehicle in response of a request



Bilateral Communication between vehicles and OEMs – How can be achieved

- The Shared Server will provide communication channels per OEM and its vehicles
- The Shared Server will utilize the **Publish/Subscribe model of communication**
- With this communication model:
 - Vehicles will send data streams with data required by the OEMs continually through a communication channel
 - OEMs will consume the data streams from the channel
 - OEMs may push notifications or updates to their vehicles by using another channel
 - Vehicles of each OEM will listen to their respective request channel and will process only the messages that are relevant to them.
 - Afterwards, they will publish a message to a response channel

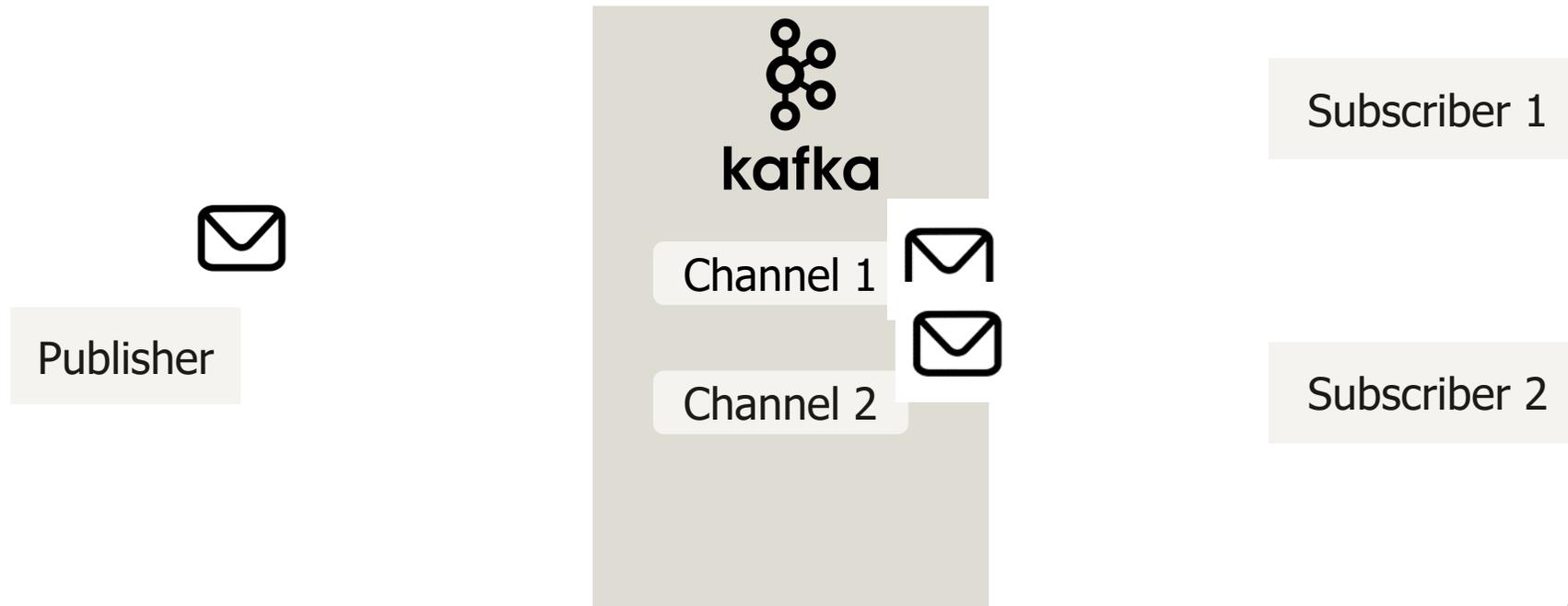
Apache Kafka

Kafka is a messaging system that is designed to be fast, scalable, fault-tolerant and durable. It is an open-source stream processing platform. Apache Kafka originated at LinkedIn and later became an open-source Apache project in 2011. It aims at providing a high-throughput, low-latency platform for handling real-time data feeds.



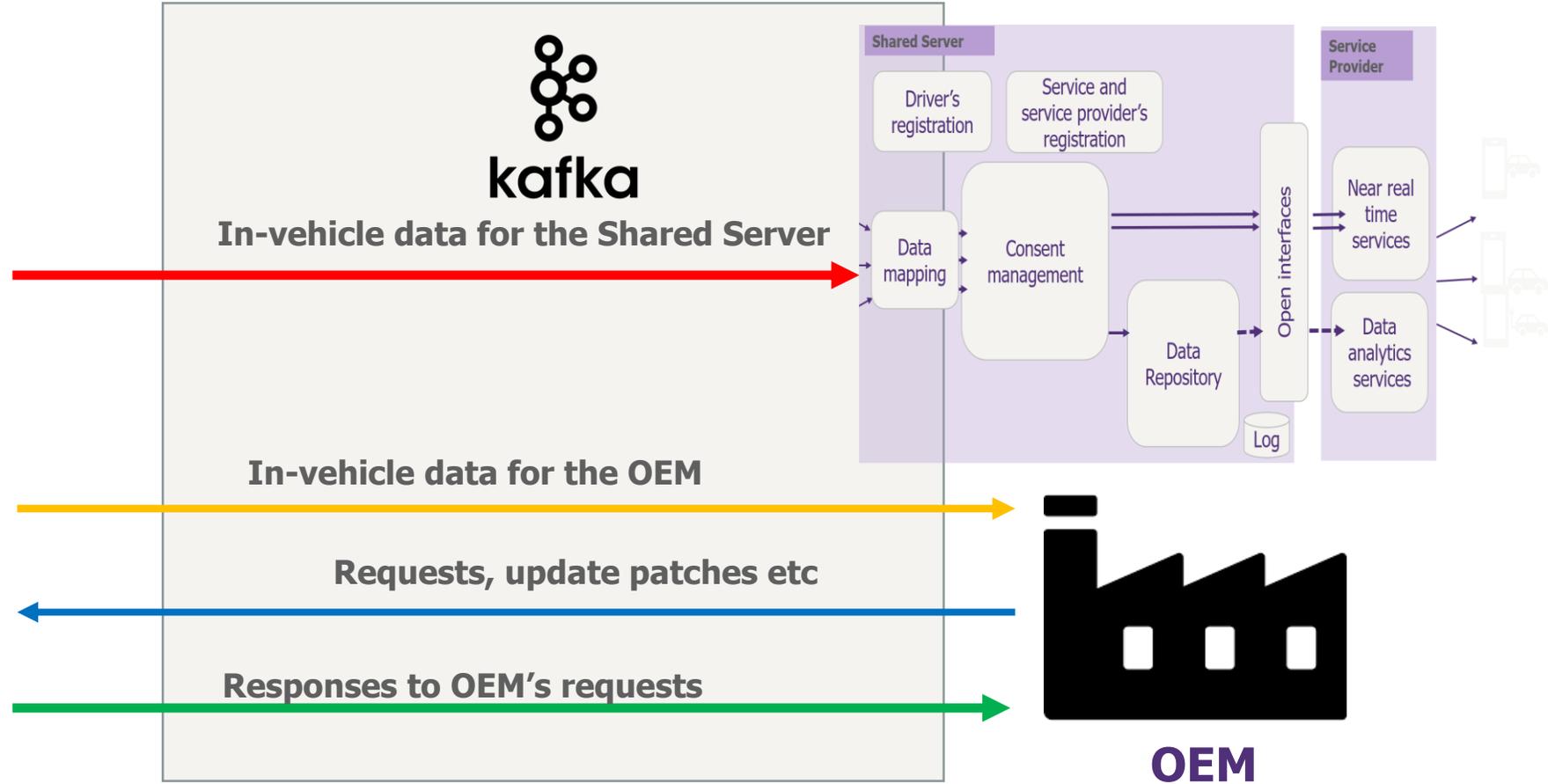
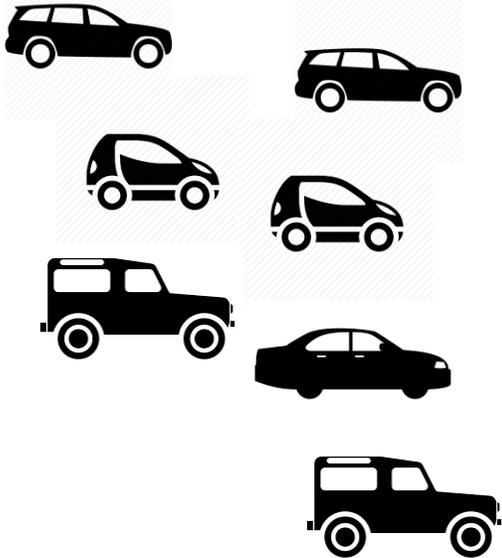


Apache Kafka provides communication channels. The applications which publish messages to these channels called Publishers. The messages are received only by the applications (Subscribers) that have subscribed to this channel.

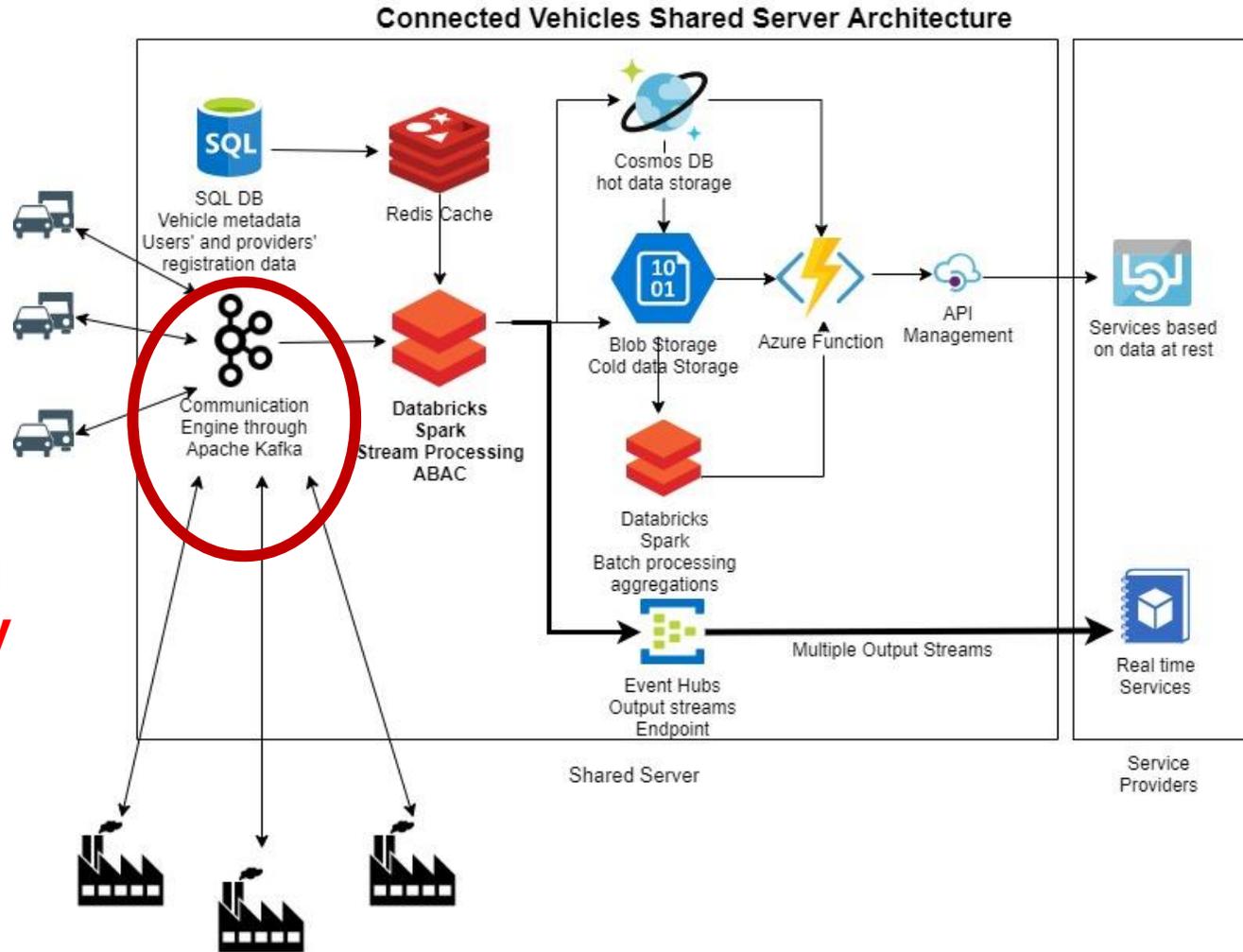


Let's take a closer look ...

Fleet of vehicles of specified OEM



Updated architecture of the Shared Server



Event hub will be replaced by Apache Kafka



Anonymised Data - **a complementary data product**

In our pilot, until now, we focused on personalised services. These services **bring great monetisation opportunity** (Hazard location Notification, Park and ride, Insurance) and the GDPR applies, asking for a challenging implementation.

The anonymization concerns the secondary use of data and should be applied **on data at rest** or cold data aiming to:

- Public authorities in order to improve traffic management and roadblocks and traffic flow data;
- **Researchers for contacting Origin-Destination (OD) of traffic** in an area of interest for a given time;
- OEMs for brand-specific applications, component analysis and p

However there is a lack of guidance on which anonymisation techniques is effective in compliance with GDPR, for data sharing on the Automotive sector.

It's a well known problem with solutions that may vary from naive approaches to more sophisticated and hard in term of computation and algorithms. Literature indicates **that data anonymisation should, therefore, defined on a case-by-case basis**, in particular in connection with the level of detail of the data (Mamoona N. Asghar et al., 2019)



Anonymised Data - **a complementary data product**

Omitting the VIN

That's is quite naive as **we lose the pos**

We also **lose the information about** information such us about the type of t exhaust emissions

However, as we may drive from home to work, mostly during working days, even without VIN, could be guessed with good probability the location of our home and work address. Then by joining this information with data from other data sources that may contain info about the home or work addresses, the user is re-identifiable.

Pseudonymisation or masking of VIN –data are still consider

In that way, the direct identification of the subject is n recurrent data tuples in the data sets, maintaining sufficient

In that sense, **Longitude and Latitude data** seem much identification of data than **oil level or temperature, indicat** information mainly regarding the mechanical status of the vehicle.

Will elaborate the case, and suggest the type of anonymization needed concerning the particular data attributes proposed for the Shared Server



Questions Answers

