AutoSPADA

Automotive Stream Processing and Distributed Analytics

Public report

Project within FFI - Automotive Stream Processing and Distributed Analytics

Author Hampus Grimmemyhr, Bastian Havers-Zulka & Ashok Koppisetty – Volvo Car Corporation
Vincenzo Gulisano, Marina Papatriantafilou, Romaric Duvignau & Elad Michael Schiller - Chalmers University of Technology
Mats Jirstrand - Fraunhofer-Chalmers Centre
Mathias Johanson - Alkit Communications AB.

Date 2023-08-25

Content

1. Summary

New automotive functions and services will increasingly rely on real-time processing and streaming of sensor data within vehicles, between vehicles, and between vehicles and cloud infrastructures, which require efficient and scalable data stream processing technology. Besides the value of the information generated through the analysis of continuous data streams from a distributed fleet of vehicles, stream processing helps in reducing the volumes of data to be transmitted, hence reducing communication time and cost overheads, as well as facilitates taking security and privacy issues into consideration. Moreover, when developing these new functions and services, connectivity, telematics, and analytics platforms are required, whereby stream processing and real-time distributed data analytics can be performed in test vehicles and selected customer vehicles, as well as in cloud services. These platforms should support distributed data analytics and data stream processing functionality in a highly flexible and configurable way, enabling the realization of many use cases with different requirements.

The aim of the AutoSPADA project has been to design, develop, and implement scalable and flexible new methods and platforms for the realization of automotive distributed data analytics and data stream processing involving both in-vehicle, near-vehicle (edge/fog) and back-end (cloud) computing resources, building on the momentum generated by the methods and proof-of-concept infrastructure previously developed in the OODIDA project. The main results of the project are novel technological concepts and software services providing a distributed data stream processing framework, demonstrating the feasibility of the approach on real vehicle fleets as well as academic results published in scientific journals and conferences.

The AutoSPADA framework is based on an architecture where processing of data streams can be performed both on-board (in vehicles) and off-board (in cloud/fog computing infrastructures) improving the scalability of data analytics frameworks, reducing vehicle communication bandwidth, while solving or mitigating many privacy concerns when transmitting unfiltered raw data from vehicles. New automotive functionalities and services based on connectivity will be enabled based on the results of the project, creating new business opportunities. The technology and concepts developed in the project furthermore have great potentials of seamlessly connecting data analysts and function developers with both the data and the computational resources available in a population of vehicles as well as in the cloud. It has been demonstrated how to facilitate rapid prototyping of new functions and services where the target applications are typically found in areas such as autonomous driving, connected active safety, remote maintenance & service, driver-assistance, battery performance analysis, and fuel economy. Also, applications requiring the aggregation and analysis of data from a population of vehicles can benefit from the developed technology and concepts, e.g., road condition monitoring. Access to real-time data and computational power of vehicle fleets in testing, verification and validation phases of product development improves product quality and shortens development cycles, reducing the time to market for new products.

The coordinator of the project is Volvo Cars. The other participants are Alkit Communications AB, Fraunhofer-Chalmers Centre and Chalmers University of Technology.

2. Sammanfattning på svenska

Nya fordonsfunktioner och tjänster kommer i allt högre grad att förlita sig på realtidsbehandling och strömning av sensordata inom fordon, mellan fordon och mellan fordon och molninfrastrukturer. Detta kräver effektiv och skalbar teknologi för strömning av data. Förutom värdet av informationen som genereras genom analys av kontinuerliga dataströmmar från en distribuerad fordonsflotta, hjälper strömning av data i att minska datavolymerna som måste överföras. Detta minskar kommunikationstid och kostnader samt underlättar hänsynstagande till säkerhets- och integritetsfrågor. Dessutom krävs anslutnings-, telematik- och analysplattformar vid utveckling av dessa nya funktioner och tjänster. Där kan dataströmning och realtidsanalyser av distribuerade data utföras i testfordon, utvalda kundfordon och molntjänster. Dessa plattformar bör stödja distribuerad dataanalys och funktionalitet för dataströmmar på ett högst flexibelt och konfigurerbart sätt, vilket möjliggör förverkligandet av många användningsfall med olika krav.

Målet med AutoSPADA-projektet har varit att designa, utveckla och implementera skalbara och flexibla metoder och plattformar för förverkligande av fordonsrelaterade distribuerade dataanalyser och strömning av data. Detta involverar både fordonsbaserad bearbetning, nära-fordon (edge/fog)-bearbetning samt bakända (moln) beräkningsresurser. Detta bygger på framdrivningen från de metoder och bevis-of-concept-infrastrukturer som tidigare utvecklats inom OODIDA-projektet. De huvudsakliga resultaten av projektet är nya teknologiska koncept och mjukvarutjänster som erbjuder ett ramverk för distribuerad strömning av data. Detta demonstrerar genomförbarheten av metoden på verkliga fordonsflottor samt akademiska resultat som har publicerats i vetenskapliga tidskrifter och konferenser.

AutoSPADA-ramverket är baserat på en arkitektur där bearbetning av dataströmmar kan utföras både ombord (i fordon) och utom bord (i moln/fogberäkningsinfrastrukturer). Detta förbättrar skalbarheten för dataanalysramverk, minskar fordonskommunikationsbandbredd och löser eller mildrar många integritetsbekymmer vid överföring av ofiltrerad rådata från fordon. Nya fordonsfunktionaliteter och tjänster som baseras på anslutning kommer att möjliggöras genom resultaten från projektet och skapa nya affärsmöjligheter. Teknologin och koncepten som utvecklats i detta projekt har även stor potential att smidigt koppla samman dataanalytiker och funktionsutvecklare med både data och beräkningsresurser som är tillgängliga i en fordonspopulation samt i molnet. Det har demonstrerats hur man kan underlätta snabb prototypframtagning av nya funktioner och tjänster där målprogrammen vanligtvis återfinns inom områden som autonom körning, ansluten aktiv säkerhet, fjärrunderhåll och -service, förarassistans, batteriprestanda-analys och bränsleekonomi. Även applikationer som kräver sammanställning och analys av data från en fordonspopulation kan dra nytta av den utvecklade teknologin och koncepten, t.ex. vägförhållandeövervakning. Tillgång till fordonsflottors realtidsdata och beräkningskraft under testning, verifiering och validering av produktutveckling förbättrar produktkvaliteten och förkortar utvecklingscykler, vilket minskar tiden till marknaden för nya produkter

3. Background

Modern vehicles contain upwards of 100 Electronic Control Units (ECUs) producing many kinds of data, such as time series signals from connected sensors, diagnostic information, images, audio, video, point clouds and more. With the advent of connectivity, data sets can be communicated over wireless infrastructures and made available for analysis for product development purposes, or for new vehicle features and services, such as connected active safety services and autonomous driving. Collection and analysis of vehicular data has a long history in the automotive industry, starting with diagnostics services being integrated in ECUs in the 1980s, enabling read-out of diagnostic data in workshops for fault tracing, quality assurance and R&D. With the advent of cellular mobile data communication networks, connected vehicles could periodically transmit diagnostic data to back-end processing infrastructure for analytics, enabling knowledge-driven product development. As wireless communication capacity increased, larger volumes of data could be transmitted, including not only diagnostic data, but increasingly also time-series signals from in-vehicle communication buses and more complex sensor data such as radar, lidar and video. Today, Automotive OEMs collect vast amounts of data from large numbers of connected vehicles, feeding data-driven analytics and statistical processing systems. The trend is towards higher volumes, lower latencies and more complex data sets, and the emergence of 5G network infrastructures, promising higher bandwidth and lower latency, will make it possible to communicate much more data with lower delays.

When designing the next generation automotive systems and services, access to these large amounts of vehicular data is of paramount importance. Therefore, telematics services and analytics frameworks are implemented in the automotive industry, giving the opportunity to capture, communicate and analyze data for decision support in product development. Traditionally, such telematics systems are based on a store-and-forward approach, whereby data is captured and stored to disk or in solid-state memory in data logging devices during a driving cycle, for subsequent upload to cloud infrastructures triggered by some event such as ignition-off. Although systems of this kind can support sophisticated trigger conditions and can implement various filtering and data reduction techniques, they are principally designed with the basic idea of communicating all data needed for a particular analysis (or service) and then performing the data processing off-board, e.g., in a cloud infrastructure.

One approach is to use data stream processing frameworks that support real-time processing of streaming data, without requiring the data to be stored before processing. Such frameworks, called Data Stream Processing Systems/Engines, Data Stream Management Systems or Complex Event Processing systems, allow operations such as queries or computations to be performed continuously on live data streams in real time. This enables new applications, where the access to data from vehicles is delay-sensitive, e.g., for connected active safety services. Moreover, since not all raw data needed for a particular analysis must be transmitted to a cloud for processing, analytics performance can be improved by distributed computations, while network bandwidth requirements are reduced, since transmission of processed or filtered data typically requires less bandwidth. Privacy concerns and other information security issues can also be alleviated by avoiding transmission of raw data from customer vehicles.

Another approach for gaining access to vehicular data and minimizing privacy exposure is to push as much computation as possible to the vehicle. When performing analysis on the vehicular data (statistical or machine learning methods) the naive approach is to upload all data to the off-board server and perform the analysis there. However, by forcing most of the computations to be performed on-board implies that only part-results need to be transmitted to the server. Decentralized techniques such as federated learning (a term coined by Google and originally applied to make mobile phones collaboratively learn a shared model, are also suitable when training machine learning models on vehicle data without needing sensitive sensor data to be sent out of the vehicle. In the FFI project OODIDA (see more information below) an initial proof-of-concept prototype of a distributed analytics framework was developed and implemented. The framework makes it possible for a user to directly connect to a fleet of vehicles and have access to the computational capabilities on-board. Tasks and assignments can be issued to both the on-board and off-board processes and only the final results of the computational task are returned to the user.

4. Project contents – planned project scope

WP 1	Management
Leader	Volvo Cars
Other participants	n/a
Description of contents	The objective of the Management work package is to make sure the project schedule is kept and to guarantee the execution of the work plan to achieve the project goals on time and within budget constraints. Moreover, WP1 will provide administrative and financial control according to the work plan, maintain technical control over the project, and ensure the integration of activities and results of the different WPs.
Method/approach	Administration, coordination, follow-up, reporting
Delivery	Periodic progress report updated every 6 months

WP 2	Use cases definitions and requirements
Leader	Volvo Cars
Other participants	FCC, Chalmers, Alkit
Description of contents	In WP2 the use cases considered in the project will be defined and their objectives, requirements, and scope will be clearly specified. During the project the use cases will be continuously monitored and updated by stakeholders in order to ensure that they are aligned with the overall project objectives.
Method/approach	Workshops, continuous meetings
Delivery	Reports/documents describing objectives and requirements for use cases

In order to specify requirements for the streaming and distributed analytics frameworks developed within the project, clear and well-suited use cases need to be included in the project. The tasks in WP2 that facilitate that the use cases are well documented and continuously updated/revised are:

Task 2.1: Use case definition

This task ensures that the use cases included in the project are well specified and documented. Workshops with the appropriate participants will be held in order to understand the use cases and to align the participants tasks and objectives. Deliverable from this task are reports clearly defining the scope and objective of the considered use cases.

Task 2.2: Continuous requirement adaption

In order to ensure that the considered use cases are still valid and aligned with the project there will be workshops twice per year (for each use case) where the participants till iteratively update and revise the requirements and tasks.

WP 3	Distributed data analytics framework
Leader	FCC
Other participants	Volvo Cars, Chalmers, Alkit
Description of contents	In WP3 the OODIDA framework will be extended and adapted for including more complex distributed analytics computations as well as stream processing tasks. The adaption of the framework will be guided by the considered use cases from WP2 in order to demonstrate the usefulness of the system. The overall objective of the WP is to take the prototype distributed analytics framework (developed in OODIDA) to a higher TRL level (TRL-6) where the defined use cases are implemented and tested on real vehicle fleets. A comprehensive survey regarding other state-of-the-art frameworks will be performed and benchmarking between found frameworks and OODIDA will be performed.
Method/approach	State-of-the-art survey, software development, method development, hardware/software integration
Delivery	Report on state-of-the-art solutions and comparisons with OODIDA framework, Distributed analytics framework (TRL-6)

The objective of WP3 is to extend and adapt the OODIDA framework to include all capabilities required from the use cases considered in the project. In order to reach these goals the work in WP3 is divided into the following tasks:

Task 3.1: Requirement specification for framework

Discussions and workshops with stakeholders will be held in the initial parts of the project in order to align the requirements and objectives of the developed distributed analytics framework. Following the continuous use case updates from WP2 revisions of the specifications of the framework may be applied.

Task 3.2: Survey of state-of-the-art solutions and benchmarking

In this task a thorough survey regarding existing frameworks for distributed analytics will be performed. Deliverable from this task is a summary report regarding found solutions and their benefits/drawbacks as well as a comparison with the OODIDA framework.

Task 3.3: OODIDA framework extension and adaptation

The OODIDA framework will be extended and adapted according to the requirement specifications from Task 3.1. Examples of extensions that will be included are further development of active code reloading for rapid prototyping, monitoring and management mechanisms for on-board processing for load balancing, and core function library extensions. The focus of this task is on production code development (TRL-6) of the framework.

Task 3.4: Hardware/software integration

This task ensures that integration with chosen hardware at the different stakeholders is considered and implemented.

Task 3.5: Distributed analytics methodology development

This task focuses on analyzing and further developing distributed machine learning methods (e.g., federated learning) focusing on existing use case definitions. The work in this task is mainly core research in mathematical and statistical methods for distributed analytics which should be tailored for the specified use cases in the project.

WP 4	Distributed stream processing
Leader	Volvo Cars
Other participants	Chalmers
Description of contents	**VCC and Chalmers - continuation of Industrial Ph.D. project:** The ongoing industrial Ph.D. project will build on the research results of the OODIDA project, extending and improving the research focusing on the development of intra-vehicle multi-sensor data streaming analysis deployed on inexpensive single-board devices that can complement the analysis run centrally at the utility (e.g., Volvo headquarters). Research dissemination will take place through journal articles, conferences and the licentiate and Ph.D. thesis of the industrial PhD student, Bastian Havers.
Method/approach	Qualitative and quantitative research, method development and applied research.
Delivery	Published scientific results in the form of peer reviewed conference papers and journal articles and a Ph.D. thesis.

The research to be conducted in this work-package is focusing on three complementary and parallel research threads. While each parallel thread can be explored autonomously from the other, the three are interconnected and designed to bring the best outcome once addressed jointly in a holistic fashion. More concretely, these threads focus on:

1. **Task 4.1**: the continuation of the research focusing on how streaming-based applications [V3,V4] can encompass the devices through which data flows (i.e., from vehicles’ sensors to vehicles' ECUs and embedded computers up to Volvo central data centers).
2. **Task 4.2**: the study of how provenance techniques for streaming analysis can be used to automatically keep track of important raw data (e.g., data that triggers a given security- or privacy-related alert) without the need for manual intervention from data analysts.
3. **Task 4.3**: the study of how GPGPU architectures, which will be soon deployed in VCC vehicles can be leveraged to further push the analysis towards vehicles rather than data towards central data center, thus accounting for efficiency, flexibility and also for security- and privacy-related concerns.

These tasks are presented in detail in the following.

Task 4.1: Streaming data analysis and trade-offs in local computation vs. data transfer

In this task, we will continue investigating streaming applications that are able to trade off data transfer for in-vehicle computation, generalizing preliminary concepts from data clustering (e.g., DRIVEN framework [V2]) to other analysis techniques. More concretely, the initial focus will be investigating how streaming-based applications can be used to monitor the fitness of locally-deployed ML techniques (e.g., based on random forests) and to minimize data transfer between vehicles and central servers by limiting the latter only to specific conditions (e.g., when a new learning phase is needed or a certain model is deemed obsolete).

Task 4.2: Leveraging of streaming based provenance techniques.

In this task, we will investigate the use of existing streaming-based techniques such as provenance [V1] in the context of security applications, to ease the deployment, debugging and prototyping of in-vehicle applications as well as the traceability of results. Based on recent research results of the Chalmers group, existing techniques (with available algorithmic descriptions and code) could be tested directly in the scope of this project.

Task 4.3: Leveraging of GPGPU architecture for in-vehicle streaming-based analysis

In this task, we will study how streaming-based analysis can be ported to GPGPU-based architectures. Differently from most of the existing research, the goal is to define hybrid solutions that, at the same time, can leverage both the CPU as well as the GPGPU of the embedded devices to be deployed in Volvo cars, in order to maximize throughput and minimize the latency of streaming-based applications deployed therein.

WP 5	Privacy-preserving methods
Leader	Chalmers, FCC
Other participants	Volvo Cars
Description of contents	In WP5, different methodologies for ensuring privacy preservation when performing stream processing and distributed analytics on vehicle fleets will be analyzed. A literature review regarding methods applicable to the automotive industry will be performed and requirement for a future project with privacy preservation as the main focus will be formulated.
Method/approach	Literature review, requirement specification
Delivery	Report on privacy preserving methods for the automotive industry, requirement specifications for future privacy preservation project

Privacy protection specifications can sometimes require severe guarantees. Differential privacy, for example, addresses problems that arise when combining multiple query results since such combinations might reveal privacy sensitive information. In the context of differential privacy, it is imperative that the results do not violate the privacy of any individual even when combining different information sources. Sensitive data should never be directly transferred from the vehicle and as much aggregation as possible should be performed on-board in order to reduce privacy problems. In this WP the objective is to investigate the type of privacy preserving methods that are needed in an automotive setting. The tasks that are included in the WP are:

Task 5.1 Architectural support for privacy-protection with data quality trade-offs

This task will start by reviewing the literature and available privacy-preserving techniques. We aim to select the most appropriate methods and propose suitable architectural components for supporting privacy protection as well as identifying benchmarks needed for system evaluation. To that end, we will experiment with a variety of methods and analyze their limits as well as trade-offs. We expect to offer novel insights into how algorithms trade between privacy protection can guarantees and the quality of the data. With these insights, it would be possible to facilitate the design of novel algorithms that can implement the methods selected for the proposed architecture.

Task 5.2 Application-oriented privacy-protection dsf

This task will evaluate the architectural components considered in Tasks 3.1 and 3.2 via the design of new solutions that satisfy the specifications of relevant privacy-protecting applications. We expect this task to provide input for the automotive industry that needs to satisfy privacy constraints when accessing privacy-sensitive data in vehicles.

Task 5.3: Requirement specification for OODIDA framework

The aim is to extend the OODIDA framework in future projects with privacy preserving methodologies. The deliverable from this task is a requirement specification regarding the software components needed for implementing privacy preserving methods. The requirements will be driven by the use cases defined in WP2.

Task 5.4: Privacy preservation and differential privacy from distributed methods

Distributed data analytics methods (e.g., federated learning) provide some level of privacy preservation due to the fact that raw data is never transferred directly from clients. Instead, the models or aggregated of the data are transferred. However, if these methods follow the strict requirements of privacy preservation is not yet clear. This task will perform an initial analysis regarding which methods provide which levels of privacy preservation.

Task 5.5: Definition and description of future project application

The aim is to initiate an additional project where the objective is to design, develop, and implement privacy preserving methods in the automotive industry. This task ensures that objectives and work packages are defined for future application.

WP 6	**Experimentation and evaluation of technology, methods and concepts**
Leader	Alkit
Other participants	VCC, Chalmers, FCC
Description of contents	In WP4, selected concepts and technology developed in earlier WPs will be tested and evaluated. Pilots for testing and evaluating the concepts in relevant environment will be performed. The WICE automotive telematics system will be used as an experimentation and development platform, where novel data streaming and privacy preservations mechanisms will be explored.
Method/approach	Agile development, action research
Delivery	Software components and frameworks for privacy preservation and streaming in edge computing platforms

When capturing in-vehicle data from external systems, such as onboard loggers and telematics systems, data processing mechanisms for streaming data are needed. This can for instance be used by OEMs and automotive component developers in supporting development of next generation in-vehicle functions and services. Stream data processing can also be supported in edge devices using fog computing approaches. In these situations, privacy preservation mechanisms are needed compatible with the stream processing paradigm.

Task 6.1: Evaluation approach and KPIs

In Task 6.1, approaches to evaluate mechanisms for privacy preservation and data stream processing mechanisms in edge computing scenarios will be developed, including Key Performance Indicators.

Task 6.2: The WICE automotive telematics system

In this task the WICE automotive telematics system will be adapted and improved in order to provide an experimentation platform for the mechanisms to be tested and evaluated in edge computing scenarios.

Task 6.3 Platform for distributed analytics and stream processing

This task will ensure that the distributed data analytics framework from the OODIDA project is further developed. The updated system will include stream processing capabilities and privacy preserving machine learning methods natively.

Task 6.4 Hardware/software integration

This task ensures that the WICE system, the OODIDA platform, and the actual hardware are all integrated into a functional demonstrator system.

Task 6.5: Piloting

This task is devoted to piloting activities, where the developed concepts and frameworks are tested and evaluated.

WP 7	**Dissemination and demonstration**
Leader	FCC
Other participants	Volvo Cars, Chalmers, Alkit
Description of contents	In WP7, results from the project will be disseminated and demonstrated. Workshops for internal and external parties will be performed where the project results are presented. Scientific results will be published in academic journals and presented at conferences.
Method/approach	Workshops, conference participation, scientific paper writing
Delivery	Scientific papers, workshops, demonstrations

5. Purpose, research questions and method

Methodology

This section aims to clarify the methodological underpinnings of the work carried out in the project. It corresponds closely with what was planned for the project in the application.

Methodology WP 1 &WP2 - Management & Use cases definitions and requirements
Not relevant/not applicable to explain in detail.

Methodology WP3 - Distributed data analytics framework
The methodologies employed for the different tasks in WP3 have been as follows:

- Interviews with the stakeholders responsible for the different use cases in order to understand the use-case specific requirements of distributed data analytics
- Surveys of existing solutions, identifying, and reviewing similar frameworks.
- Software development through agile software development methods, with weekly internal meetings to follow the progress. Larger bi-monthly meetings with stakeholders for the different use cases.
- A high-level road-map document for the development

Methodology WP4 - Distributed Stream Processing
The activities within WP4 have been carried out by the Industrial Ph.D. student Bastian Havers-Zulka under the supervision of the academic partners. The research has been conducted, on the one hand, with other Ph.D. students, postdoctoral students and faculty members at Chalmers, while on the other in collaboration with the Industrial supervisors. Periodic meetings (on a bi-monthly basis) have allowed academic and industrial supervisors to synchronize and plan together the following steps. Thanks to the joint research between Chalmers and Volvo Cars, the research being conducted could be tested and evaluated with data and systems from Volvo Cars.

Methodology WP5 - Privacy-preserving methods

The objective of WP5 was to assess the inherent ability of the solution provided in OODIDA Phase I to preserve privacy. Through extensive investigation, we identified several well-known attacks that these solutions are unable to address effectively. Additionally, we conducted a thorough review of the state-of-the-art literature in privacy preservation, with a specific focus on the automotive domain. This comprehensive analysis and literature review were presented in the form of a detailed survey report (D5.1).

During the early stages of the project, we reviewed methods and existing applications related to privacy preservation in the automotive context. Deliverable 5.1 provided comprehensive insights into privacy-protecting solutions that are relevant to the automotive industry as a whole.

Methodology WP6 - Demonstration and experimentation

In WP6 the methodological approach, particularly in the earlier phase, included experimentation activities, wherein an iterative and agile development process was linked with tests of novel concepts and newly developed software, with direct feedback to the development.

Methodology WP7 - Dissemination and demonstration

Dissemination and demonstrations of the work conducted in the project ocurred both in academic as in industrial settings via peer-review of scientific papers and hands-on demonstration sessions.

Originality and newsworthiness

Data analytics in general, and data stream processing in particular, is currently a very hot research topic in many application areas. Connected vehicles and applications enabling advanced driver assistance, connected active safety services and autonomous driving features are top priorities for all automotive OEMs.

The uniqueness and main news value of the project lies in the innovative combination of state-of-the-art distributed analytics, data stream processing, and cloud computing technology from the ICT industry with telematics and analytics technology from the automotive industry. From an international perspective, Volvo Cars has a strong position with regards to connected services and data-driven product development. To keep this leading position, and to advance the state-of-the-art in customer data collection and analytics, the results of the AutoSPADA project have been important in both seeding new developments but also concrete products such as a new compression implementation for telematics (see WP6 Results). With connectivity being a central feature of next generation vehicles, telematics services have gained a renaissance and is currently a very hot topic in the automotive industry and in the automotive technology research field. The opportunity of doing large scale data stream processing both in onboard and in cloud computing environments further strengthens this trend. Whereas automotive security and privacy have been addressed in several previous FFI-funded projects (cf. e.g., HEAVENS, HoliSec, CASUS, ThreatMove, CyReV), the combined focus of stream data processing and privacy is novel, and in particular the question of how-to tradeoff between data quality and privacy has proven a fertile research topic.

The knowledge created in the project has the potential to affect the automotive development process in a profound way. Connectivity in combination with distributed analytics and data stream processing, including both onboard and cloud-based data analytics, will have a significant impact on the automotive industry as a whole during the next few years, with completely new stakeholders from the ICT industry entering the scene. This project has contributed to maintaining Sweden's leading position both in automotive technology development and ICT development (as demonstrated by publications in the most renowned conferences of the field, see for example *Ananke* in Disseminations and Publications).

The project was built on concepts and knowledge gained during the precursory OODIDA project (cf [OODIDA final report, 2020]). In that project, a proof-of-concept framework was developed together with analysis of novel methodologies for stream processing and distributed machine learning. The AutoSPADA project has focused more on use-case-driven research where the objective has been to adapt and further extend the distributed analytics framework to actual real-world cases from the automotive industry.

6. Objective

The objectives of the AutoSPADA project have been described in detail in the application and have been used as a gnomon throughout the project. For ease of reference, they are summarized in the following paragraphs.

A main objective has been to design, develop, and implement systems for automotive data stream processing and distributed data analytics building on the results in the previous FFI-project OODIDA (cf [OODIDA final report, 2020]) (associating also with earlier ones such as BAuD II). The systems should involve both on-board processing, back-end (cloud) processing and edge device processing (fog computing) that preserves data quality and privacy based on the need from the data consumer. The final distributed analytics framework can seamlessly connect data analysts with both the data and computational resources available in a population of vehicles and the cloud, as well as it can facilitate rapid prototyping of new functions and services utilizing on-board vehicle data processed by algorithms running both on-board and off-board on central servers. Applications are in areas such as autonomous driving, safety, maintenance & service, driver-assistance, and battery and fuel economy. Also, applications requiring the aggregation and analysis of data from a population of vehicles can benefit from the developed framework (e.g., road condition monitoring). Moreover, users of the developed distributed analytics framework need not necessarily be limited to in-house OEM data analysts but can potentially also offer tasks to 3rd party developers facilitating new business models.

The distributed analytics and stream processing architecture and methods developed in the project target to improve the quality of next-generation products, to contribute to reducing the time to market for new products and services, and to give competitive advantages by having the best available technology for connectivity, data capture and analytics. These high-level goals are met through the following sub-goals:

- Building of knowledge about distributed analytics and data stream processing in automotive contexts
- Exploration of different stream processing technologies and distributed computing approaches
- Improve stream processing methods for complex automotive sensor data, including lidar, radar and video
- Exploration into how the orthogonal requirements of powerful high-quality data capture and end-user privacy can be addressed in a systematic way
- Development and implementation of a distributed analytics framework seamlessly connects data analysts with both the data and computational resources available in a population of vehicles as well as in the cloud. The work has been up to TRL-6 for the framework, with future plans for commercialization of the software.
- Extensions of a state-of-the-art automotive telematics system with novel data streaming mechanisms, in order to test and experiment with data stream processing in test vehicle fleets
- Establishment of a position at the forefront of research and innovation in the field by producing high-quality academic results, disseminated through conferences and scientific publications

The project contributes to the following overarching FFI objectives:

- **Increasing research and innovation capacity in Sweden, thereby securing competitiveness of Sweden's automotive industry**

- **To create internationally connected and competitive research and innovation environments in Sweden**

- **Promoting the participation of small and medium-sized enterprises**
Alkit Communications AB is an SME.
- **Promoting cross-sectoral cooperation**
Promoting cooperation between industry, universities, and higher education institutions

- **Undertake development initiatives of relevance to industry**

- **Ensure that new knowledge is developed and implemented, and that existing knowledge is implemented in industrial applications**

The project contributes to the FFI subprogramme *Electronics, software and communication* objectives mainly within the two subareas *Architecture* and *Intelligent and reliable systems.*

Architecture
The project is aligned with the objectives of the *Architecture* subarea of the *Electronics, software and communication* program. With more and more sensors being available in the vehicles the need for scalable and efficient data stream processing methods is very important, and this project designed and developed methods that concern both the internal communication in the vehicles (process streams of data on-board) as well as the external communication with cloud-based servers. Computational power, sensor data, machine learning, data analysis and shorter lead times enabled by rapid prototyping are among the domains where the AutoSPADA project contributes.

Intelligent and reliable systems
A large part when extending and adapting the analytics framework has been to further develop and implement various distributed machine learning and high-dimensional statistical analysis methods, which makes the AutoSPADA project aligned with the purpose of the *Intelligent and reliable systems* subarea in the *Electronics, software and communication program*. The specific trends highlighted in the subarea that are important for AutoSPADA are: *Complex vehicle functions and systems of systems, Individualized functionality, Data-driven development of products and services, as well as novel logging of field data (*lidar, radar, video)*, through the WICE automotive telematics system.*

7. Results and deliverables

AutoSPADA has been an explorative, demand-driven research and development project with a clear emphasis on improving industrial competitiveness and innovation in the automotive industry. In this way we believe it has contributed strongly to the FFI program's expressed goals and with the sub-program's scope. On a high level, the results from the project can be summarized as:

Developed concepts and technology, including the AutoSPADA framework (targeting TRL-6), dimensioned and implemented according to the use case requirements. The developed technology has been deployed and validated in real vehicle fleets, constituting the base for future commercialized products.
Knowledge increases at all project partners within data stream processing and distributed analytics
Definition and proposal of future application within distributed analytics where the focus is on privacy-preserving methodologies.
A doctoral thesis with focus on data stream processing (Bastian Havers-Zulka).
Proof-of-Concept demonstrations of the implemented use cases.
In-depth project results and achievements are presented per work package below.

WP3 Results (Distributed data analytics framework)

The objective of WP3 was to extend and adapt the OODIDA framework in order to lift the prototype implementation to TRL-6 (*Technology demonstrated in relevant environment*), which has been accomplished by the deployment and demonstration of the AutoSPADA distributed analytics system on two Volvo C40 during spring 2023. The demonstrations were performed on two different occasions with use cases denoted *One Pedal Drive* and *Durability Analysis*, respectively. Prior to these demonstrations in a real industrial environment, simulated validation and verification use cases have been performed both internally at FCC and at AI Sweden’s Edge Learning Lab such as *Ice Spot Detection and Mapping* and *Distributed Active Learning*.

The work package consisted of three main tasks: T3.1 Requirement specification for framework, T3.2 Survey of state-of-the-art solutions and benchmarking, and T3.3 OODIDA framework extension and adaptation. A summary of the results from each task is given below and a bit more elaborate description of the AutoSPADA software platform as well as brief use case descriptions follow.

T3.1 Requirement specification for framework
Due to the pandemic no initial workshops to kick off the work in T3.1 were arranged. However, by having discussions and digital meetings first in the project group and later with other Volvo Cars stakeholders (e.g., autonomous drive, security, people experience, data protection) requirement specifications (D3.1) were gradually stated. Furthermore, updates and add-ons to the specification of the framework were continuously included during the project.

T3.2 Survey of state-of-the-art solutions and benchmarking
An extensive review of suitable software architectures for the framework has been performed. This was summarized in a report (D3.2) outlining specifications and comparisons regarding scalability, reliability, resource constraints, security and privacy, and development budget.

T3.3 OODIDA framework extension and adaptation *(FCC AutoSPADA platform)*
This task encompasses the design and development of the AutoSPADA edge computing platform for distributed data analytics developed by the Fraunhofer-Chalmers Centre (FCC).

Modern vehicles produce large amounts of data, which are highly interesting for product & function tests and verification as well as for product development. Transferring this data to a central server becomes infeasible when the number of involved vehicles and recorded signals grows. It also prevents real-time processing. The FCC AutoSPADA platform addresses both problems and aims at seamlessly connecting data analysts working with high-level tools, such as Python notebooks, with both the data and computational resources available in a population of vehicles as well as in the cloud. It has been designed for real-time distributed big data analytics, targeting connected vehicles.

The platform facilitates rapid prototyping of new functions and services utilizing on-board vehicle data processed by algorithms running on-board, which are also easily integrated with algorithms running off-board on central servers, a typical example being distributed learning approaches such as federated learning. AutoSPADA has been deployed on the WICE system utilizing the Edge Learning Lab facilities hosted by AI Sweden and has also been deployed on Volvo Cars test vehicles during the final stages of this FFI project. Applications are expected to be found in areas such as

- Safety
- Maintenance & service
- Driver-assistance
- Fuel economy
- A/B testing
- Distributed statistical databases
- Onboard HD map tuning
- Monitoring battery health
- Active data collection

Also, the development of applications requiring the aggregation and analysis of data from a population of vehicles would benefit from AutoSPADA, e.g., road condition monitoring. Below follows a more detailed description of the platform, its requirement specification, technology, architecture, and development process (D3.3-5).

Overview

The AutoSPADA computing platform is the successor to the OODIDA computing platform developed by FCC in the previous FFI project *BADA – On-board Off-board Distributed Data Analytics* (OODIDA) and founded on the lessons learned through the OODIDA project. Although the concept is similar, the platform features a completely redesigned architecture to allow for a higher technology readiness level (TRL). Compared to other platforms in the edge computing space, AutoSPADA has a strong focus on privacy and security, while imposing very few limitations on the types of tasks that can be assigned to edge devices.

The purpose of the AutoSPADA platform is to allow for rapid development and deployment of user-provided computational tasks on edge devices, such as on-board microcontroller units (MCUs) in e.g., automotive vehicles. In this project, the client code has been deployed on Host Mobility's MX-4 units in collaboration with Alkit Communications as part of their in-vehicle WICE platform. The AutoSPADA platform is designed to support a wide range of task paradigms, such as e.g., data aggregation, monitoring, or machine learning including paradigms like federated learning. Results from clients are collected by the server infrastructure and can be received in real-time through streaming channels or retrieved on demand. An important concern has been to enforce privacy and security throughout the platform, to prevent attacks by a third party, or even from malicious tasks. As the platform is intended to be deployable on a vast number of clients, a scalable design has also been a major architectural concern.

The AutoSPADA platform is designed to support virtually any task implementation language, with Python being the currently supported choice. Hence, the current mode of operation for users of the AutoSPADA platform such as data scientists and engineers developing distributed analytic tasks is to design client code and server code using high-level Python programming.

AutoSPADA uses a proven IoT oriented protocol with a minimal network footprint for notifications. For updates and data transfers, a binary remote procedure call (RPC) framework is used, where server and client APIs are compiled from a common interface specification to minimize protocol mismatches between communicating parties.

AutoSPADA has a strong focus on privacy and security. All network communication is secured by state-of-the-art encryption, communicating parties are always mutually authenticated, and tasks are isolated from their hosts through containerization.

A major consideration in the design of the AutoSPADA platform is maintainability. Important parts of this aspect are easy access to a rich ecosystem of official third-party libraries and a language with low barriers of entry (Python).

Platform Design

The distributed AutoSPADA platform is built around three classes of nodes. *Client nodes* are responsible for spawning tasks on demand and reporting the results of these tasks. *User nodes* submit tasks to be run on the clients, and can also retrieve task results, either streaming, or on demand. The *server nodes* bridge the gap between the client and user nodes, by receiving task requests from users and results from clients and persisting and/or forwarding these. This organization mirrors the roles of the actors in the platform, as illustrated in Figure 1.

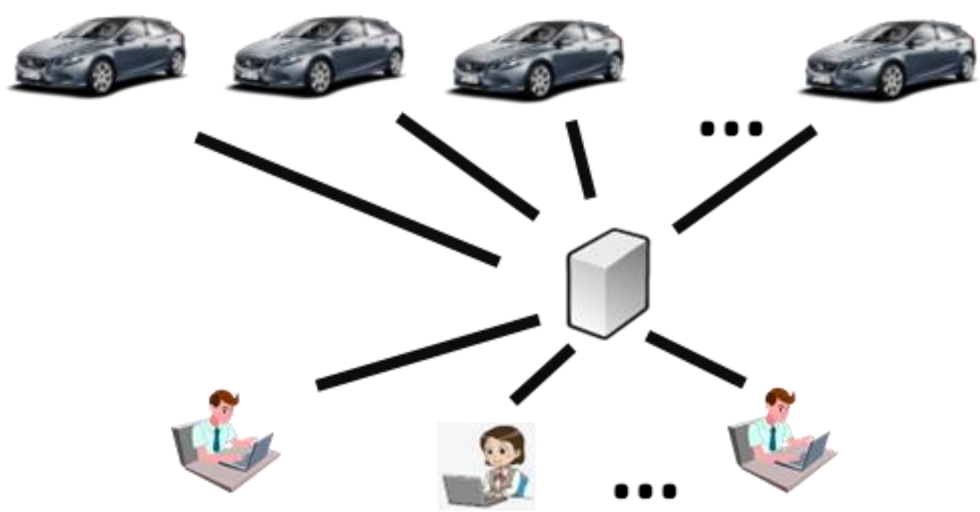

Figure 1. The logical AutoSPADA actors are the users (bottom) and the clients (top). The infrastructure of transmitting tasks and results from the users to the clients and vice versa, respectively, is implemented by the server nodes (middle).

Platform Architecture

The overall architecture of the AutoSPADA platform is illustrated in Figure 2.

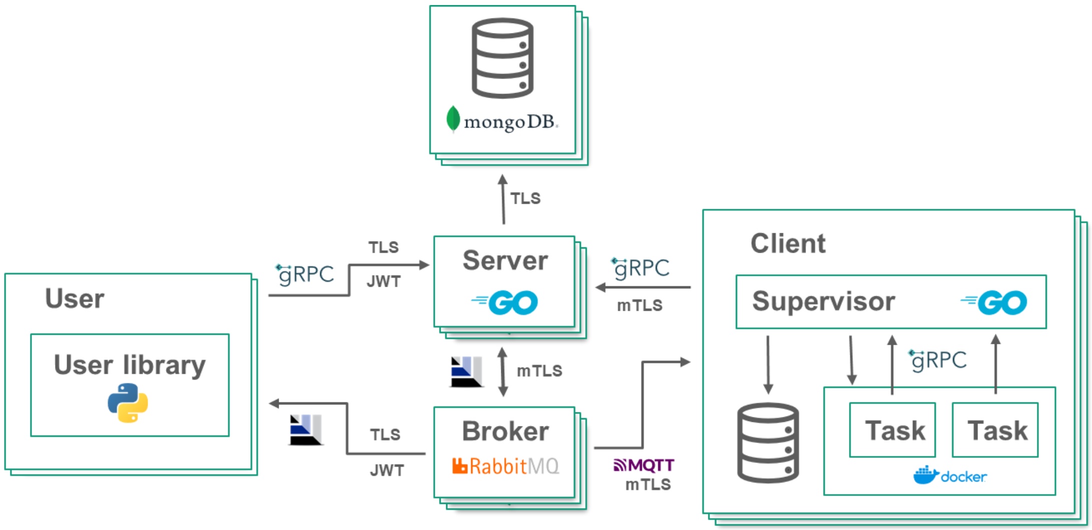

Figure 2. An overview of the frameworks, languages and protocols used in the AutoSPADA platform. All communication is secured by TLS. Authentication between nodes is performed by OIDC using JSON Web Tokens (JWTs) or by mutual TLS (mTLS) using X.509 certificates.

Deployment, Validation, and Verification process

During the different development stages the AutoSPADA platform has been deployed in different ways, see Figure 3, utilizing both in-house resources, commercial services, and AI Sweden's Edge Learning Lab facilities.

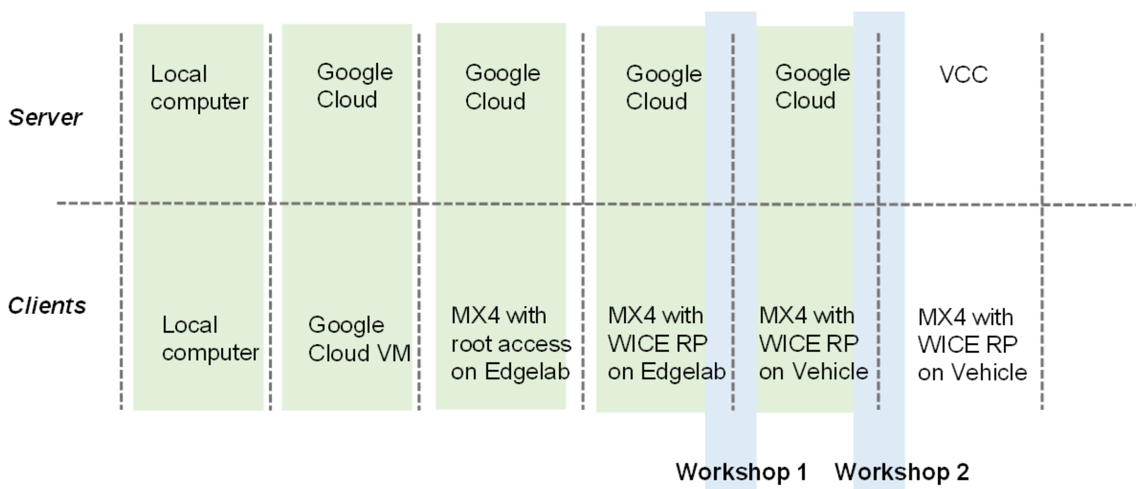

Figure 3. The deployment, validation, and verification process of the AutoSPADA platform. Client and server code are deployed from local to cloud to rack-mounted hardware to in-vehicle hardware (green – accomplished, white - future). The two dissemination workshops close to the end of the project are marked in blue.

Use Cases

One Pedal Drive (OPD) is a setting in modern electric vehicles that, when active, brakes whenever the acceleration pedal is released, meaning that a driver typically does not have to use the brake pedal. The automatic breaking in OPD also regenerates energy to the battery. However, it is more energy efficient to reduce speed by coasting (i.e., driving with OPD Off) instead of breaking and regenerating with OPD turned On.

This use case aimed to investigate what impact OPD has on energy consumption for different drivers by having them drive a test route with and without OPD. To measure this, several Key Performance Indicators (KPIs) can be calculated, which include the time spent in OPD mode, brake pedal use, acceleration pedal use, total energy consumption, and the amount of regenerated energy.

Durability of vehicle hardware depends on how it will be used. Certain maneuvers can result in excessive wear and introduce heavy loads on, for example, steering components. It is therefore important to identify user behaviors to extrapolate how that will affect the durability of a unit.

This use case investigated driver behaviors in terms of steering wheel angles in relation to the speed of the vehicle. The behavior is encoded into two-dimensional heatmaps, as visualized in Figure 4. A hypothesis is that extreme angles will result in more wear and be exaggerated by the speed. If the hypothesis holds and can be mapped to the behaviors, a heatmap computed from real-time data can then be used as a prediction of durability.

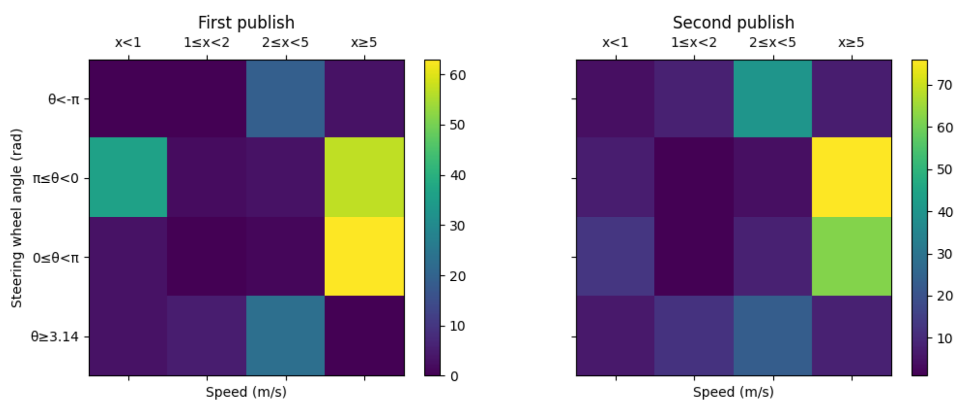

Figure 4. A heat map showing the number of observations for the combination of speed and steering wheel angle from two separate data collections.

Ice Spot Detection and Mapping is a use case that demonstrates the potential of a distributed platform in continuously updating a database with real-time events from vehicles. The primary objective of this use case is to develop a product iteratively that can enhance road safety by providing up-to-date information about slippery road conditions. By leveraging their own slippage sensors as ground truth, vehicles contribute to the constant refinement of a map indicating areas with slippery or icy surfaces. This collaborative effort ensures that when a vehicle encounters a newly formed ice spot, the database map is promptly updated, benefiting the entire fleet.

To accomplish this, a simulation environment was employed, wherein each vehicle was equipped with pre-recorded sensor values generated from a framework called simulation of urban mobility (SUMO) with randomly placed ice spots. These ice spots periodically emerged and melted away over time. As a vehicle traverses through one of these spots, a comparison between the recorded map and the real-time data is conducted to trigger appropriate updates. Figure 5 visually depicts a snapshot of the ice spot database.

One of the key challenges lies in determining the most fault-tolerant approach to updating the database, which is where AutoSPADA can ease the development of a practical and functional solution. Additionally, the use case demonstrates data integration of an external database that is independent of the platform. Likewise, other services like public weather data could also be utilized.

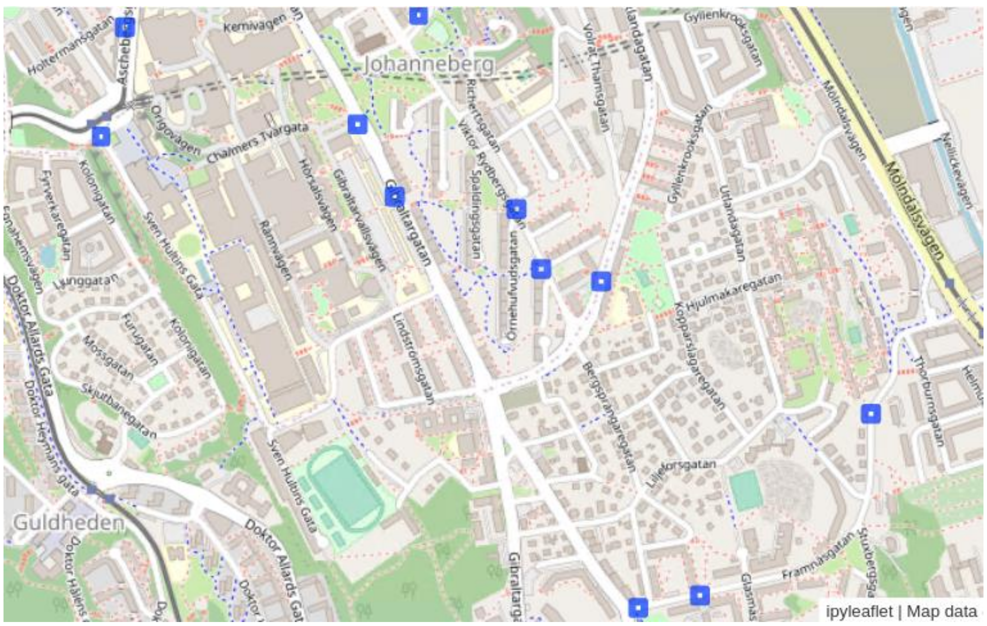

Figure 5. A geographical map representation of the database containing the ice spots at a specific point in time.

Distributed Active Learning is a machine learning research field that aims to develop algorithms to identify the most informative data samples for human annotation. In the distributed setting, data is sampled locally on distributed clients. These algorithms work by quantifying the current model's uncertainty about unlabeled samples. Hence, an Active Learning (AL) algorithm avoids collecting data that the model already handles well while simultaneously serving as a method to better capture rare occurrences. However, since the choice of algorithm is problem-specific, they have to be evaluated for the problem at hand. This evaluation process translates well into the rapid iterative development enabled by AutoSPADA.

We developed a use case to demonstrate an AL development flow as well as show how AutoSPADA can be used in combination with familiar machine learning tools and libraries such as PyTorch and TensorBoard. The use case evaluated the so-called Least Confidence AL algorithm and a random sampling baseline on the classification of MNIST images using a convolutional neural network model. Image data was distributed among a small number of simulated AutoSPADA clients. From the results in Figure 6, we can see that the Least Confidence algorithm consistently outperformed random sampling, which could motivate further evaluation that includes more AL algorithms. Working iteratively and exploratively in this way, whether it is evaluating new algorithms or fine-tuning a parameter, is a key motivator behind the platform.

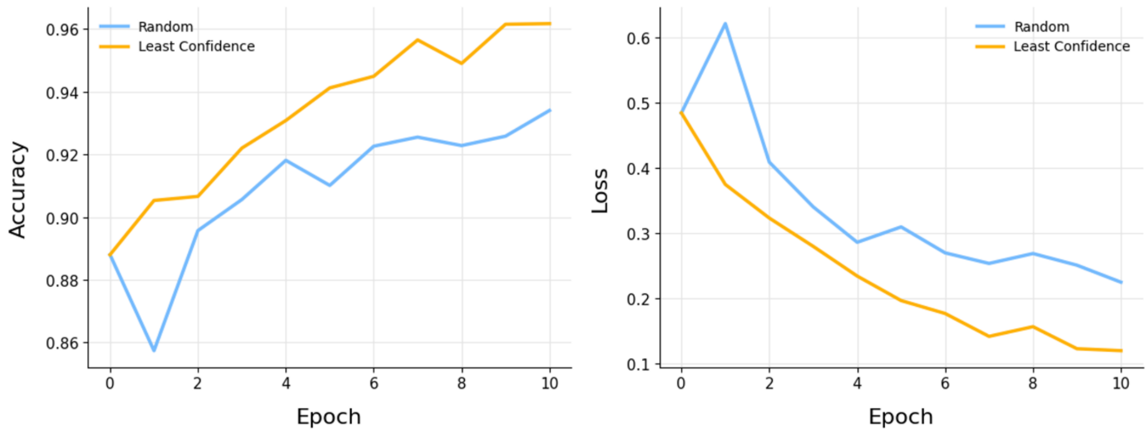

Figure 6. Comparisons of two models trained on data sampled using the Least Confidence algorithm and random sampling, respectively. The diagrams show evaluation with respect to accuracy and loss, respectively, on a test set. To be comparable, we made sure that both algorithms started from the same initial model and training data, which is also why they share a point at the zeroth epoch.

WP4 Results

WP4's research focused on distributed and communication-efficient data analysis in Vehicular Cyber-Physical Systems (VCPS) consisting of connected vehicles. Such vehicles continuously sense increasingly large data volumes from high-bandwidth sensors such as LiDAR and can leverage the computational power of edge devices on-board vehicles and of a central server while balancing bandwidth and performance. In the following, outcomes are grouped based on the resulting research publication.

A) DRIVEN

This research thread, started in the previous VINNOVA project called OODIDA, focused on the implementation and evaluation of a streaming algorithm for lossy data compression and its embedding into a data gathering and analysis framework in the streaming domain. The resulting framework combines lossy data compression and data clustering in a data pipeline and was evaluated under simulation of different network conditions and different tolerable bounds for the compression loss. The results were published in a conference and a journal article (cf. Relevant chapters and references in [Havers Lic 2020]), and part of the framework is being used in actual vehicles and in the WICE platform (See this report, Section 6, WP6 Results).

B) Data localization at vehicular networks' edge
This research thread has investigated smart query dissemination in a vehicular network with respect to analysis cost and response time, simulating query answering on vehicles through embedded devices called ODROID. Among other aspects, the research focused on fairness aspects of the proposed query dissemination algorithms as well as adapted these algorithms to a fleet that continuously loses and adds vehicles [Duvigau et al 2021].

C) Ananke, forward provenance
The aim of this research thread was to extend a streaming backward provenance framework to support low-overhead delivery of a novel concept we introduced under the name of live forward provenance [Palyvos et al 2021]. With this work, we have shown that our proposed framework, called Ananke, can be deployed on low-powered devices such as those found in vehicles, to perform for example intelligent data selection on the edge. We have also bult on Ananke, to draw explanations about missing results in stream processing, implemented in a framework called Erebus [Palyvos et al 2023]. Both these works appear in highly influential venues (presentations in VLDB conference instances and PVLDB journal). Further, there is increased interest in the methods, as they lead to novel, improved approaches for data selection and management, via continued research efforts.

D) Distributed learning strategies in vehicular Cyber-Physical Systems
This work's focus is to devise the requirements for a framework that can evaluate various strategies for distributing the process of learning from data in a system consisting of connected vehicles equipped with on-board computers, road-side units, and a central coordinating server. The framework simulates message exchange between the system actors via long-range and short-range communication, enabling analysts to investigate the effects of offloading mobile data transfer to low-cost direct vehicle-to-vehicle or vehicle-to-road-side-unit communication on the performance of learning. A prototype implementation has been used to evaluate a distributed Machine Learning (ML) strategy that could not have been assessed using conventional ML frameworks [Havers et al 2022].

E) Nona – dynamic forward provenance
In contrast to the work on forward provenance in Ananke, this work seeks to devise dynamic forward provenance for data streaming queries that are dynamically deployed and removed during the runtime of the provenance system to offer vastly enhanced flexibility. This research thread extends the definitions of forward provenance to the dynamic case, develops a deeper understanding of the evolving nature of provenance, and devises ways of always guaranteeing the properties of forward provenance during the dynamic evolution of the system (defined by the addition and removal of queries) [work to be submitted for publication in Q4 2023].

WP4 defined a total of 6 deliverables:

1. **[D4.1]** Publication focusing on data streaming provenance in the context of vehicular networks (in connection with Task 4.1)
2. **[D4.2]** Licentiate Seminar for Bastian Havers-Zulka (Industrial Ph.D.)
3. **[D4.3]** Extended journal publication on data streaming provenance in the context of vehicular networks (Task 4.1)
4. **[D4.4]** Publication about trade-offs in data analysis and communication within the context of vehicular ML techniques (Task 4.2)

5. **[D4.5]** Publication about the use of GPGPU architectures in the context of vehicular streaming-based analysis (Task 4.3)
6. **[D4.7]** Ph.D. Defense for Bastian Havers-Zulka (Industrial Ph.D.)

The results achieved within WP4 align with those defined by the initial deliverables, with only a small deviation, as explained next. Deliverable **D4.1** has been completed in connection with the work on live forward provenance (C), which also resulted in a publication at the A* conference VLDB (see Publications in Section 7.2). Deliverable **D4.2** was completed on the 10th of November 2020 (see Chapter 7.2, Licentiate Thesis). Deliverable **D4.3** has been completed in connection with both the work on live forward provenance (C) and that on data localization at vehicular networks' edge (B). Deliverable **D4.4** has been completed within the research activities about distributed learning strategies in vehicular cyber-physical systems (D) and resulted in a publication. A small deviation concerns deliverable **D4.5**, which has been merged within **D4.4**, since the latter focus on ML frameworks that also support learning tasks on GPGPU architectures. Deliverable **D4.7** is expected to be completed in November 2023, according to the extension in the studies for the industrial Ph.D. student Bastian Havers-Zulka, due to parental leave. Additional results will result in a publication focusing on the topic of dynamic forward provenance (E).

WP5 Results

Building upon the findings of the survey report, we formulated recommendations to enhance OODIDA's solutions in terms of privacy preservation, with a particular emphasis on the application of differential privacy (Task 5.3). By adopting differential privacy techniques, we propose to use existing architectural solutions for the development of an architecture specifically tailored to the automotive context. Such architecture aims to facilitate online machine learning applications without the necessity of transferring privacy-sensitive information, thus ensuring the protection of individual privacy.

To illustrate the potential use of the proposed application, we present it in Task 5.4. This section showcases the future applications and benefits of the privacy-preserving architecture, focusing on two specific scenarios: unnecessary injuries prevention and driver's distraction warning. By leveraging the proposed architecture and privacy-preserving techniques, we aim to address critical challenges while safeguarding individuals' privacy.

Overall, the work carried out in WP5 involved an analysis of OODIDA's privacy preservation capabilities, exploration of well-known attacks, a comprehensive literature review, and the formulation of recommendations for improvement. The integration of differential privacy techniques and the development of a privacy-preserving architecture hold great promise in advancing OODIDA's solutions and ensuring privacy protection in the automotive domain.

Task 5.4: Future Applications of the Proposed IMS Privacy-Preserving Architecture

We propose a privacy-preserving architecture for learning from the Interior Monitoring System (IMS) by opening avenues for various future applications in the automotive context. By ensuring the protection of privacy-sensitive information, the IMS can be leveraged to facilitate online machine learning applications without compromising the privacy of individuals. This section outlines potential future applications and benefits of the proposed architecture.

1. Unnecessary Injuries Prevention:

The IMS can play a crucial role in preventing unnecessary injuries by establishing a connection between one's posture and potential injuries. By continuously monitoring the driver's posture and providing real-time feedback, the system can promote safe driving practices. This includes reinforcing the importance of fastening seatbelts to prevent injuries caused by airbag inflation. Studies have highlighted the risks associated with airbag deployment, such as major injuries resulting from improper positioning or objects obstructing the airbag's path. By leveraging online data and privacy-protected learning, the IMS can identify additional hazards and contribute to the development of improved airbag systems that minimize the risk of unnecessary injuries.

2. Driver's Distraction Warning:

Driver distraction poses a significant risk on the roads, contributing to a considerable number of accidents and fatalities each year. The IMS can enhance existing driver distraction warning systems by providing clearer and more accurate warnings to drivers. By leveraging various sensors and advanced algorithms, the IMS can detect common sources of driver distraction, including daydreaming, cell phone use, outside events, and other occupants. The IMS can then issue timely alerts and reminders to help drivers maintain focus on the road. With approximately 30,000 fatalities per year in the USA attributed to distracted driving and a substantial portion of crashes being caused by driver distraction, the IMS has the potential to significantly reduce these numbers.

By integrating the IMS with the proposed privacy-preserving architecture, these future applications can be developed while safeguarding the privacy of individuals. The architecture's key features, such as summarization and anonymization of important features, as well as the utilization of avatar-like sketches of the passengers and differential privacy to further protect their privacy. This is to ensure that privacy-sensitive information is not transferred or compromised.

Overall, the IMS with the privacy-preserving architecture represents a pioneering solution in the automotive context, setting a precedent for online machine learning applications without the need for sharing privacy-sensitive information. The future applications discussed above demonstrate the potential benefits of this approach in promoting safety, reducing injuries, and enhancing driver awareness on the road.

The work in WP5 resulted in two additional publications that further contribute to the understanding and development of privacy preservation techniques. In the publication by Dolev et al. 2023, the focus was on privacy protection using erasure codes for distributed storage systems. This research explored how erasure codes can be utilized to enhance privacy and security in distributed storage environments. Additionally, the publication by Georgiou et al. 2020 delved into the broader architectural context, specifically addressing fault tolerance and edge computing. By examining these aspects, the research aimed to identify potential synergies between privacy preservation, fault tolerance mechanisms, and edge computing paradigms. These publications extend the knowledge and possibilities within the field of privacy preservation and serve as valuable resources for OODIDA's ongoing development.

WP6 Results

The work in WP6 has concerned the use of the WICE Automotive Telematics System for experimentation with the novel software developments of the AutoSPADA project, and to support demonstrators. The WICE system itself was improved in the project by introduction of novel functionality (e.g., the PLA algorithm as described below, and extensions to the Rapid Prototyping framework).

WICE as demonstration and experimentation platform

In tasks 6.2 and 6.3 of WP6 (Experimentation and testing of technology, methods and concepts), the WICE automotive telematics system, developed by Alkit, was adapted and improved in to provide an experimentation platform in the AutoSPADA project. This also included hardware/software integration, in providing an in-vehicle hardware platform to execute the AutoSPADA software services developed by FCC. In-vehicle signals were made available to experimental software implementations using a publish/subscribe API known as the WICE Signal Broker. The Rapid Prototyping module of WICE (developed in previous FFI project BAuD II) was heavily used to deploy software in connected vehicle fleets.

A pilot use case for experimentation by Volvo Cars engineers in collaboration with FCC and Alkit was conducted, and several demonstrations were performed.

PLA implementation in WICE

The PLA (Piecewise Linear Approximation) algorithm has been implemented in the WICE Signal Reader software module (see Figure 7). This work has been part of WP6: Experimentation and testing of technology, methods and concepts) and has been carried out by Alkit Communications in collaboration with Volvo Cars. The PLA algorithm itself was developed in WP4 - Distributed Stream Processing in collaboration between Volvo Cars and Chalmers.

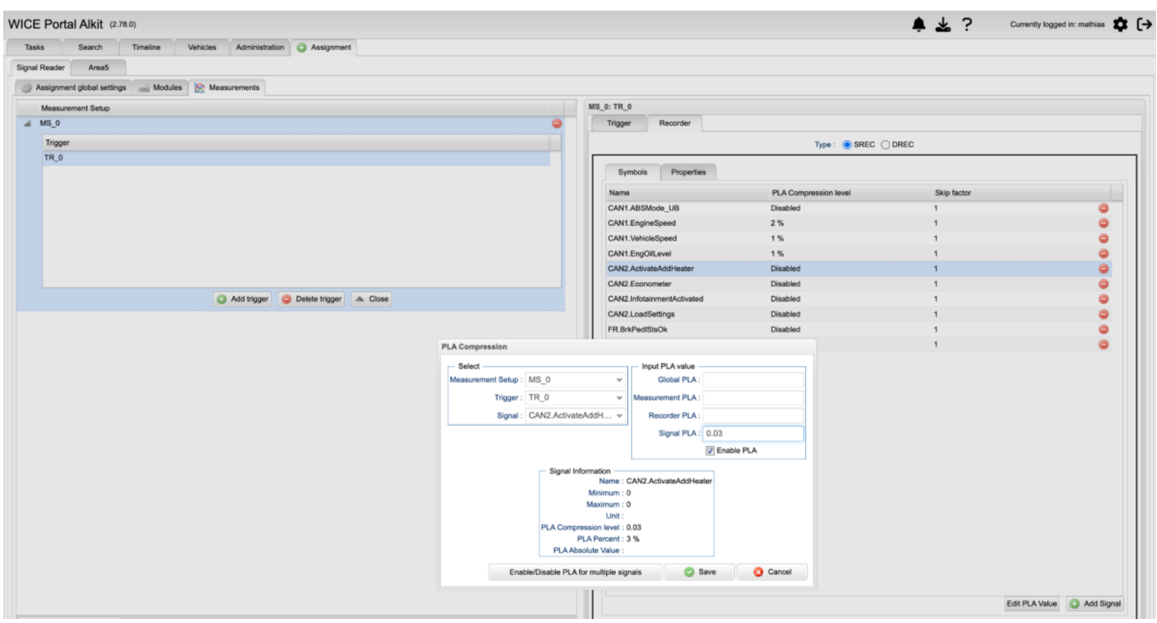

Figure 7: Screenshot of the Signal Reader editor in WICE where measurement assignments for in-vehicle signal data collection is configured. The screenshot shows how PLA compression can be enabled for specific signals and the compression level selected.

Experiments with the PLA implementation has been performed in the WICE-equipped vehicle fleet at Volvo Cars and in specific test vehicles at Alkit. Some early usage experiences indicate that PLA can reduce data volume significantly, but compared to other approaches the gain is not so big (e.g., “on-change” sampling, zip compression). Since end-users are not used to working with approximate data, they generally prefer on-change sampling to PLA, but given that they gain more insight into the complete data collection process, they can be expected to adapt to working with lossy compression of signals. For specific use cases, where precision of the data collected is to of foremost interest, it can be very useful.

Figure 8 shows an example of signals collected using PLA sampling. When zoomed-in, the signal plots exhibit the linear approximation which is the central mechanism of data reduction in the algorithm.

Figure 8: Screenshot of visualization of signals collected using PLA sampling. The line segments between the points are the linear approximations, where multiple data points would have been present using conventional sampling.

8. Dissemination and publications

8.1 Dissemination

How are the project results planned to be used and disseminated?	Mark with X	Comment
Increase knowledge in the field	X	
Be passed on to other advanced technological development projects	X	
Be passed on to product development projects	X	
Introduced on the market		
Used in investigations / regulatory / licensing / political decisions	X	

8.2 Publications

8.2.1 Conference publications

WP4 [Havers et al 2022] *Proposing a Framework for Evaluating Learning Strategies in Vehicular CPSs.* Bastian Havers, Marina Papatriantafilou, Ashok Koppisetty and Vincenzo Gulisano. ACM/IFIP International Middleware Conference. 2022.
https://doi.org/10.1145/3564695.3564775

WP4 [Palyvos et al 2022] *Research Summary: Deterministic, Explainable and Efficient Stream Processing.* Dimitris Palyvos-Giannas, Marina Papatriantafilou and Vincenzo Gulisano. Advanced tools, programming languages, and Platforms for Implementing and Evaluating algorithms for Distributed systems (ApPLIED) Workshop, Held in conjunction with PODC. 2022.
https://doi.org/10.1145/3524053.3542750

WP5 [Georgiou et. al 2020] *A Self-stabilizing Control Plane for Fog Ecosystems.* Zacharias Georgiou, Chryssis Georgiou, George Pallis, Elad Michael Schiller, Demetris Trihinas. UCC 2020: 13-22. 2020.
https://doi.org/10.1109/UCC48980.2020.00021

8.2.2 Journal articles

WP4 [Gulisano et al 2023] *An Algorithm for Tunable Memory Compression of Time-Based Windows for Stream Aggregates*. Vincenzo Gulisano. International Workshop on Scalable Compute Continuum (WSCC 2023), co-located with the 29th International European Conference on Parallel and Distributed Computing (Euro-Par 2023). 2023.

WP4 [Palyvos et al 2023] *Erebus: Explaining the Outputs of Data Streaming Queries.* Dimitris Palyvos-Giannas, Katerina Tzompanaki, Marina Papatriantafilou, Vincenzo Gulisano. 49th International Conference on Very Large Data Bases (VLDB). VLDB Endowment, Vol.16, No.2. 2023.
https://dx.doi.org/10.14778/3565816.3565825

WP4 [Duvignau et al] *Time- and Computation-Efficient Data Localization at Vehicular Networks' Edge.* Romaric Duvignau, Bastian Havers, Vincenzo Gulisano and Marina Papatriantafilou. IEEE Access 9. 2021.
https://doi.org/10.1109/ACCESS.2021.3118596

WP4 [Palyvos et al 2021**]** *Ananke: A Streaming Framework for Live Forward Provenance.* Dimitris Palyvos-Giannas, Bastian Havers, Marina Papatriantafilou, Vincenzo Gulisano. 47th International Conference on Very Large Data Bases (VLDB) / ,Vol.1 No. 4. 2021..
https://doi.org/10.14778/3430915.3430928

WP5 [Dolev et al 2023]. *Self-Stabilizing and Private Distributed Shared Atomic Memory in Seldomly Fair Message Passing Networks.* Shlomi Dolev, Thomas Petig, Elad Michael Schiller. Algorithmica 85(1): 216-276. 2023
https://doi.org/10.1007/s00453-022-01023-w

WP5 [Zhang et al. 2022] *Evaluation of Open-Source Tools for Differential Privacy.* Zhang, Shiliang, Anton Hagermalm, Sanjin Slavnic, Elad Michael Schiller, and Magnus Almgren. Sensors 23, no. 14: 6509. 2023.
https://doi.org/10.3390/s23146509

8.2.3 Masters Theses

WP4 Assim Sulaiman Khaled**.** A framework for smart vehicular data-collection. [in collaboration with Volvo Cars]. 2023. In progress.

WP4 Erlandsson, Andréas, and Mikael Gordani Shahri. 2021. "Interactive Fine-Grained Provenance for Streaming-Based Analysis Applications."
https://search.ebscohost.com/login.aspx?direct=true&db=ir01625a&AN=cst.20.500.12380.302287&site=eds-live&scope=site.

WP4 Holmgren, Jakob, and Philip Nord. "Evaluating the Benefits of Spatio-Temporal Relational Operations for Validating LiDAR Perception Systems," 2021.
https://search.ebscohost.com/login.aspx?direct=true&db=ir01625a&AN=cst.20.500.12380.304103&site=eds-live&scope=site.

8.2.4 Other references

WP4 [Havers Lic 2020] *Distributed and Communication-Efficient Continuous Data Processing in Vehicular Cyber-Physical Systems.* Bastian Havers. Licentiate Thesis. 2020.
https://research.chalmers.se/publication/519767/file/519767_Fulltext.pdf

[OODIDA final report, 2020] *BADA – On-board Offboard Distributed Data Analytics.* Max Peterson. Final project report within FFI – Big Automotive Data Analytics (BADA), 2020. https://www.vinnova.se/globalassets/mikrosajter/ffi/dokument/slutrapporter-ffi/effektiva-och-uppkopplade-transporter-rapporter/2016-04260eng.pdf

9. Conclusions and future research

Throughout this research project, several complementary research threads were explored, towards the goal of improving localization and value extraction from large volumes of distributed data sensed by vehicular fleets. Each research thread not only advanced the state of the art but also laid the groundwork for future steps, both within the same research thread and through merging and complementing the various research threads.

The results achieved during this project are significant and impactful. They can be summarized as follows:

- Distributed algorithms for data localization in Vehicular Networks were developed. These algorithms are designed to select vehicles that triggered specific conditions or events, aiming to balance the time required to identify a subset of vehicles with relevant data and the computational overhead incurred by each vehicle in checking the validity of its data against a set of properties.
- Support for better debugging and information gathering was addressed through several avenues:
    - Richer provenance information was introduced. This information not only specifies which source data contributes to specific data analysis results but also indicates whether each piece of source/raw data can potentially contribute to future results. This distinction aids analysts in prioritizing the inspection of the vast volumes of events observed during the monitoring of Cyber-Physical Systems (CPSs), particularly in connection with privacy-sensitive analysis.
    - A novel formal definition of the problem of explaining missing answers in streaming applications was proposed. Furthermore, a framework was introduced that allows users to validate and debug streaming queries by defining boolean expectation predicates on the query outputs. This framework, called Erebus, verifies whether expected results are produced and provides explanations for the absence of expected results (missing answers).
    - A tool was developed to evaluate various learning strategies in Vehicular Cyber-Physical Systems. This tool assists fleet operators and original equipment manufacturers (OEMs) in learning from data generated by the vehicles themselves in a manner that is optimal for their specific fleet and use case.
- Experiments with onboard data stream processing for pilot Use Cases in Volvo's test vehicle fleet, using the WICE automotive telematics platform and the AutoSPADA framework.
- Development and implementation of new data sampling algorithm (PLA) in WICE.

- A greatly enhanced FCC AutoSPADA platform for distributed data analysis (successor of the FCC OODIDA platform), which aims at seamlessly connecting data analysts working with high-level tools, such as Python notebooks, with both the data and computational resources available in a population of vehicles as well as in the cloud. Close to the end of the project the platform reached TRL-6.
- The FCC AutoSPADA platform facilitates rapid prototyping of new functions and services utilizing on-board vehicle data processed by algorithms running on-board, which are also easily integrated with algorithms running off-board on central servers.
- Several use cases demonstrating the use of the FCC AutoSPADA platform for distributed data analytics (one pedal drive analysis, durability analysis, ice spot detection and mapping, and distributed active learning).
- Preparations for follow-up projects and exploration of new business possibilities based on novel technology and new concepts developed in the project.
- The project results also entail proposing a novel privacy-preserving automotive application, developed through an in-depth analysis of existing literature and utilizing state-of-the-art methodologies.

The individual and holistic significance of these results not only opens avenues for continued and new research but also holds the potential to drive value creation, particularly in collaboration with Volvo. Potential directions for further exploration and collaboration include:

- Investigating how to adapt the proposed techniques to Vehicle-to-Vehicle (V2V) scenarios, exploring the portability of the developed algorithms.
- Advancing the development of finer-grained debugging and knowledge-extraction models and tools. This includes refining provenance information and exploring techniques for smart data selection to enhance the value extraction processes. Notably, accounting for elastic analysis, as currently being explored by researchers from Chalmers' team, could provide valuable insights in this regard.
- Exploration of possibilities with onboard data stream processing and Edge Computing for Federated Learning applications.
- Investigation of legal aspects of onboard data stream processing in global customer car fleets.

The FCC AutoSPADA platform’s current technology readiness level would make it very easy to rapidly set up, deploy, and perform medium to large scale distributed data analytics use cases on vehicles equipped with WICE, which would be a natural target for future projects involving current partners in this project as well as other automotive partners in the FFI program.

10. Participating parties and contact persons

Academic partners:

Chalmers University of Technology, Department of Computer Science and Engineering (CSE). Contact person: Vincenzo Gulisano, Biträdande Professor (vincenzo.gulisano@chalmers.se). Other faculty/students involved: Elad Michael Schiller (Biträdande Professor), Marina Papatriantafilou (Associate Professor), Romaric Duvignau (Assistant Professor), and Bastian Havers-Zulka (Industrial Ph.D. Student).

Fraunhofer-Chalmers Centre, Department of Systems and Data Analysis (SYS). Contact person: Mats Jirstrand, Head of Department (mats.jirstrand@fcc.chalmers.se). Other contributing staff: Emil Gustavsson (Business Area Leader - Machine Learning and AI), Magnus Önnheim (Core Competence – Machine Learning and Optimization), Jonas Hagmar (Senior Developer), Simon Smith (Senior Developer), Adrian Nilsson (Senior Developer).

Industry:

Volvo Car Corporation Ashok Koppisetty (ashok.chaitanya.koppisetty@volvocars.com)
Volvo Car Corporation Hampus Grimmemyhr (hampus.grimmemyhr@volvocars.com)
Volvo Car Corporation Bastian Havers-Zulka (bastian.havers@volvocars.com)

Industrial partner:

Alkit Communications AB. Contact person: Mathias Johanson, CEO (mathias@alkit.se). Other contributing staff: Mangus Fant (Senior Developer), Jonas Jalminger (Senior Developer). Johan Strand (Senior Developer), Jonathan Olsson (Junior Developer)