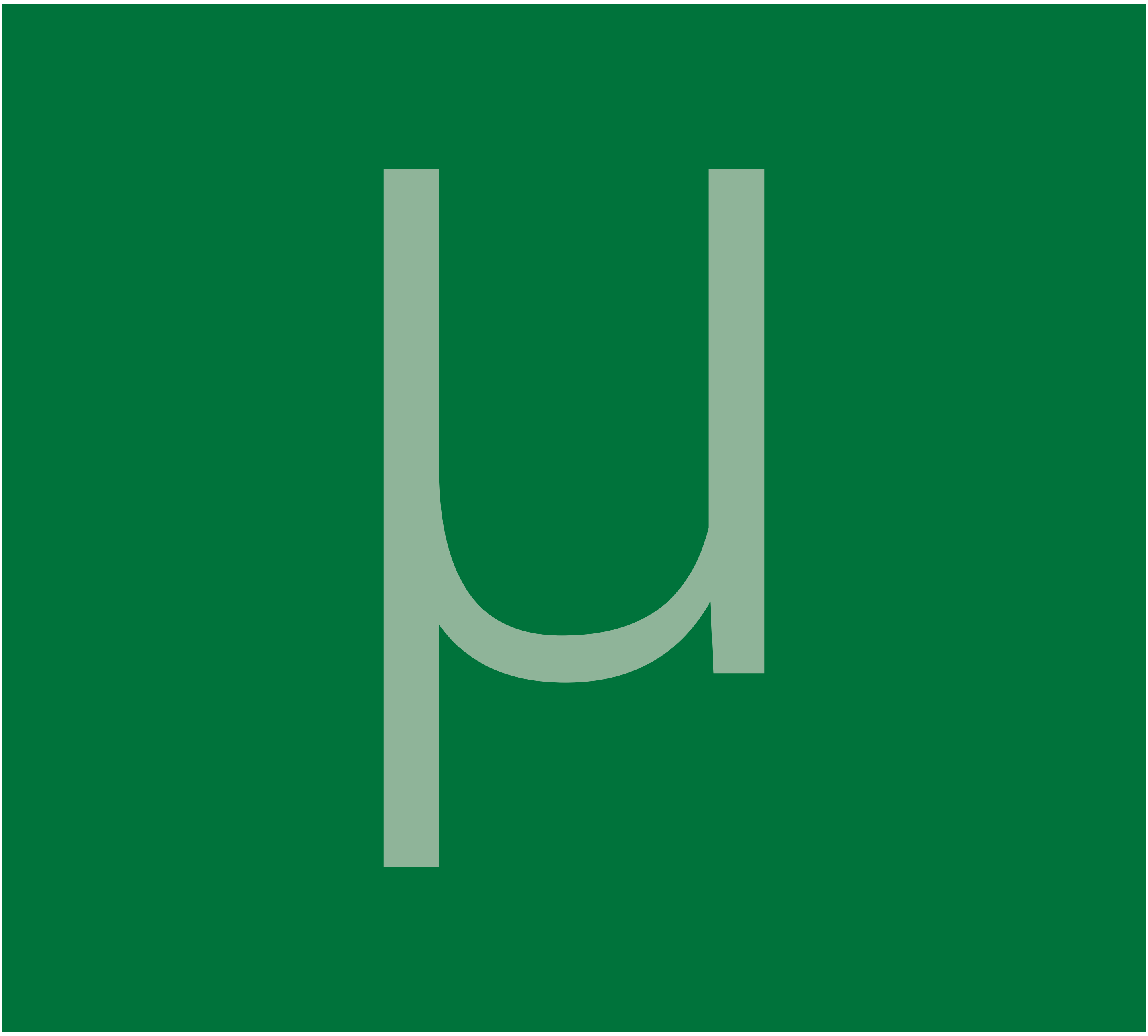
VDA Verband der
Automobilindustrie

FAT Forschungsvereinigung
Automobiltechnik

FAT-Schriftenreihe 371

Study on the technical evaluation of decentralization based de-identification procedures for personal data in the automotive sector



Study on the technical evaluation of decentralization based de-identification procedures for personal data in the automotive sector

Forschungsstelle

Goethe Universität Frankfurt am Main
Chair of Mobile Business & Multilateral Security
Institute of Business Informatics

Autoren

Prof. Dr. Kai Rannenberg
Dr. Sebastian Pape
Frédéric Tronnier
Sascha Löbner

Supported by Christian Gartner and Sabina Aliyeva

Das Forschungsprojekt wurde mit Mitteln der Forschungsvereinigung Automobiltechnik e.V. (FAT) gefördert.

Executive Summary

The aim of this study is the technical evaluation of de-identification methods based on decentralization, in particular methods of distributed and federated learning for personal data in concrete use cases in the mobility domain. The General Data Protection Regulation (GDPR) has significantly increased the incentive and effort for companies to process personal data in compliance with the law. This includes the creation, distribution, storage and deletion of personal data. Non-compliance with the GDPR and other legislation now poses a significant financial risk to companies that work with personal data. With a substancial increase in computing power at the users' side, distributed and federated learning techniques provide a promising path for de-identification of personal data. Such methods and techniques enable organizations to store and process sensitive user data locally. To do so, a sub-model of the main model that processes data is stored in the local environment of the users. Since only necessary updates are transmitted between the submodel and the main model, two advantages can be achieved from this approach. First, there is no central database, which makes it immensely difficult for potential attackers to obtain large amounts of data. Second, only fragments of the locally stored data are transferred to the main model. In the first work package of this report, suitable use cases for this study are identified through a scientific literature review. The following use cases are identified and analyzed with regard to data, benefits, model and sensible data: Traffic flow prediction, Energy demand prediction, Eco-routing, Autonomous driving, Vehicular object detection, Parking space estimation.

In the second work package, attack scenarios and general countermeasures against these attacks are discussed. To do so, relevant transmission paths, data types and trust scenarios are considered. On the one hand, it can be seen that Federated Learning has a high potential to improve the communication between entities in different scenarios and thus to also improve the accuracy and usability of the applications. On the other hand, we find that Federated Learning is not a standalone privacy preserving machine learning technique and needs to be combined with other techniques. In the last work package, countermeasures that can be used in combination with Federated Learning are discussed. A designated test network is created to evaluate the potentially achievable level of data protection for the identified use cases.

Contents

1 WP1 - Identifying Use Cases

The aim of WP1 is to identify use-cases that have the potential to benefit from decentralized/federated learning. Thereby, we build upon the results of the previous project [FAT, 2021] in which several potentially suitable use cases were identified. In this work package we additionally focus on the identification of new use cases that benefit from additional privacy protection, or such use cases that cannot be implemented using conventional privacy-preserving methods. The objective of this work package is twofold: Firstly, to conduct a literature review to identify existing use cases in which federated learning is implemented and to assess which criteria and requirements need to be met to ensure applicability of federated/decentralized learning. Secondly, to check whether and how the identified solutions can be transferred to use cases relevant to this work. In the beginning we set the following requirements for the use cases. First, the use case has to be implemented in a vehicle scenario or close to it. This means, entities involved should be vehicles or entities directly communicating with vehicles. Second, the use case should be realistic. Keeping in mind that we aim to build a demonstrator in WP3 the use cases are required to be likely to be implemented. Moreover, for realistic use cases the probability is higher that data or previous research exists. Third, the use case needs to be computable. With limited resources on the edge and in the vehicle, the use case should be able to run on devices with limited computational capacity.

1.1 Literature Review

To obtain an extensive overview of existing federated learning models in the vehicular domain we investigated relevant literature from scopus and google scholar. Thereby, we identified 62 scientific publication that are listed in table 1.2.6 in the appendix. The table lists only paper which include federated learning and a vehicle-related topic in the abstract or description.

1.2 Results

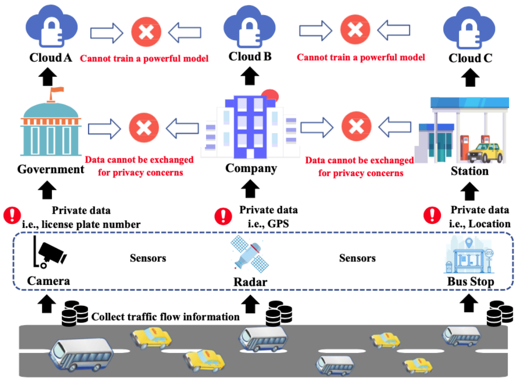
In this section we take a closer look at the existing use cases we found in the literature review. We provide a brief description of the use case. Moreover, we identify the benefit of federated learning for the respective use case.

1.2.1 Traffic flow prediction

Interested parties in traffic flow prediction are manifold and include all types of vehicle users, e.g., taxi drivers or parcel delivery services.

Although models for traffic flow prediction already exist, Liu et al. [2020] and Xu and Mao [2020] claim that traffic flow prediction with federated learning can highly profit from the decentralized learning because data sharing issues can be overcome. The data sharing issues are illustrated in figure 1.1. Liu et al. [2020] also provide the example of traffic flow cameras, that cannot easily share data because the collected pictures contain sensitive information e.g. vehicle license plates (VLP).

Figure 1.1 – Fuel optimization by route choice [Liu et al., 2020]


Cloud A
Cannot train a powerful model
Cloud B
Cannot train a powerful model
Cloud C
Government
Data cannot be exchanged for privacy concerns
Company
Data cannot be exchanged for privacy concerns
Station
Private data i.e., license plate number
Private data i.e., GPS
Private data i.e., Location
Camera
Sensors
Radar
Sensors
Bus Stop
Collect traffic flow information

With federated learning, models can become more accurate and more up to date because no anonymization at a central B-IP is required. At the same time, the data of participating vehicles (horizontal federated learning) is kept secret and with vertical federated learning, different institutions can share their data.

Table 1.1 – Traffic Flow Prediction

Data	Intersections, Streetcameras, Smart Sensors, Radar
Benefits	Higher accuracy, more up to date models because data can be shared among entities
Model	Traffic monitoring/prediction, Route predition
Sensible Data	GPS, VIN/ID, Faces

1.2.2 Energy Demand Prediction

Saputra et al. [2019] propose a faderated learning based energy demand prediction scheme that is utilizying the charging station data (see 1.2). The charging stations are clustered based on their energy demand prediction in high and low demand areas. Although this topic is highly relevant for the change in transportation, taking more sustainable solutions into account, we will not further consider this use case because it does not exhibit on board vehicle data.

Figure 1.2 – Energy demand prediction [Saputra et al., 2019]

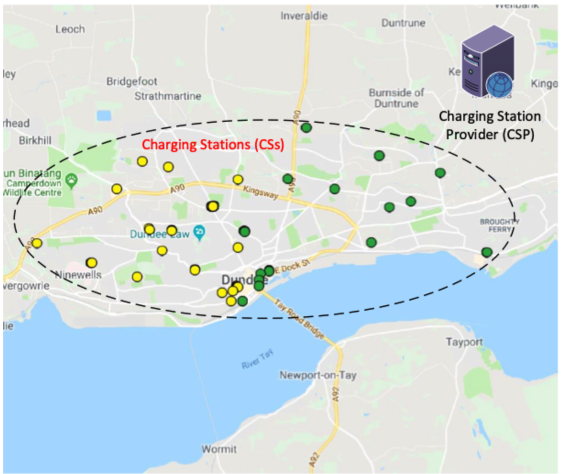


Table 1.2 – Energy demand prediction

Data	Location of CSs in Dundee City, UK, 2017 and 2018
Model	Energy demand prediction of next period (high/low region cluster)
Benefit	Better energy planning for EGO
Sensible Data	Log files with all transactions of EVs charging (CS ID, EV ID)

1.2.3 Eco-routing

Oh et al. [2019] present a dataset collected from 383 vehicles in Ann Arbor, Michigan, USA. The data includes time-series data such as speed, energy, fuel and auxiliary power usage. They also include different types of vehicles that are 27 PHEV/EVs 92 HEVs and 264 gasoline vehicle. Figure 1.3 and figure 1.3 show how different routes can effect time and fuel consumption.

Figure 1.3 – Fuel optimization by route choice [Oh et al., 2019]

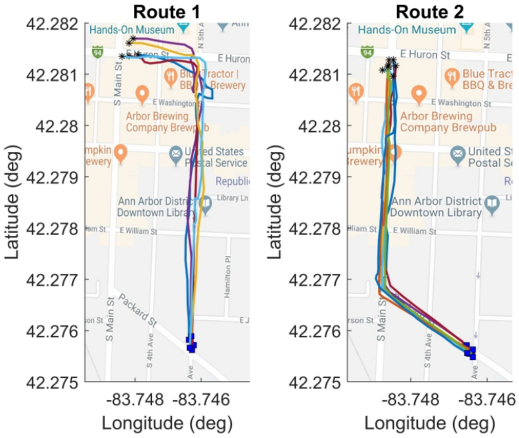
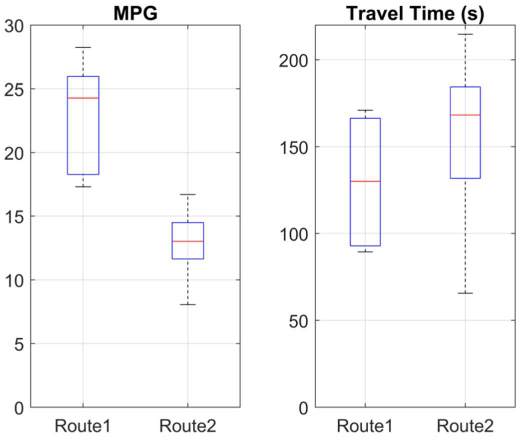


Figure 1.4 – Different fuel consumption by route Oh et al. [2019]



With cities trying to reduce their $CO2$ emissions, an energy consumption optimized route planning can help to reduce $CO2$. The model could also be optimized in a way that areas that exhibit a very high $CO2$ value are bypassed.

This scenario could well profit from a federated learning network. A model could be trained with the routes driven. The central server collects the weights for a route prediction model that optimizes the fuel consumption.

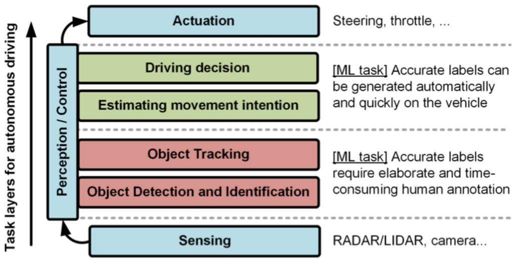
Table 1.3 – Eco-routing

Data	383 vehicles in Ann Arbor, Michigan, USA
Benefits	Reduce/manage C02 emission in cities
Model	Route prediction with optimized fuel consumption
Sensible Data	GPS, VIN/ID

1.2.4 Autonomous Driving

Autonomous driving is one of the key technologies for future traffic concepts. But rather than a single tasks, autonomous driving combines a variety of different machine learning applications. Elbir and Coleri [2020] have classified these applications that are shown in figure 1.5.

Figure 1.5 – Autonomous driving tasks for machine learning Elbir and Coleri [2020]


Task layers for autonomous driving
Perception / Control
Actuation
Steering, throttle, ...
Driving decision
Estimating movement intention
[ML task] Accurate labels can be generated automatically and quickly on the vehicle
Object Tracking
Object Detection and Identification
[ML task] Accurate labels require elaborate and time-consuming human annotation
Sensing
RADAR/LIDAR, camera...

When it comes to constantly updated machine learning applications, federated learning is one of the key technologies. Especially for classification tasks, such as object detection and identification, federated learning can be used. Some applications will be explained in more detail in the following.

1.2.5 Vehicular Object Detection

In general, there is a huge amount of object detection tasks that are often used in the vehicular environment [Wang et al., 2021]. This can range from simple traffic signs to the detection of clothes people are wearing. The most prominent examples that were identified in the literature review are presented in the following:

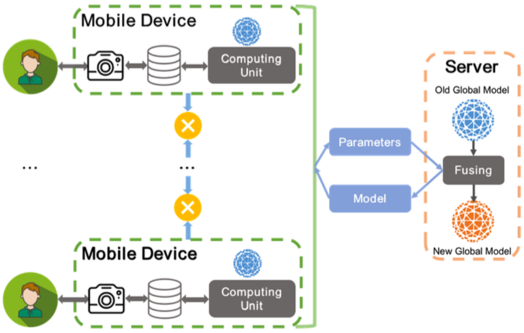
Traffic Sign Classification

One machine learning application that often is related to driving is the classification of traffic signs. While this problem has already be tackled in literature, a federated learning based framework as e.g. proposed by Nuding and Mayer [2020] is new.

Licence Plate Identification

The recognition of license plate is a very good example of highly sensible data in the public area. Although the license number is public available, tracing and profiling are likely if licence plate data is merged from different sources. This will open the door to further attacks. Kong et al. [2021] claim that one of the main use cases of licence plate identification are with the traffic authorities. Examples are identification of parking violators. Up to them, especially in China, road cameras are used to spot such violations. But also with a police officer making photos with a mobile device, a reliable and trustworthy identification of license plate numbers is important.

Figure 1.6 – License plate recognition [Kong et al., 2021]



The main challenges in this field, are management and control of transport, privacy issues and the requirement of large resources for computation, especially in big cities.

Table 1.4 – Vehicular Object Detection

Benefit	Learn objects from other countries
Sensible Data	GPS, Face, Walk

Federated learning can help to connect isolated islands of data and increase the privacy of each vehicle driver/owner by keeping the licence plate information locally. Although a lot of data about the domestic licence plates might be available, foreign license plates might be a problem that can be overcome with federated learning.

1.2.6 Parking Space Estimation

Space estimation when parking a vehicle is used in mostly every up to date vehicle. But especially complicated situations that are not contained in the original training data are hard to train. Huang et al. [2021a] propose a shared LSTM model for parking space estimation. Their model structure is shown in figure 1.7 , Lu et al. [2021]

Figure 1.7 – Parking Huang et al. [2021a]

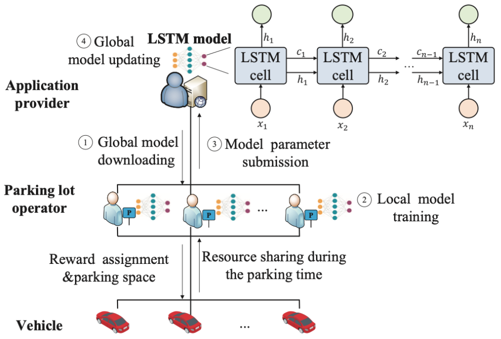


Table 1.5 – Parking Space Estimation

Data	Vehicle and parking lot sensors
Benefits	Learn complicated parking situations
Model	Parking lot operator
Sensible Data	Camera data, sensor data, location

In this scenario, federated learning has two advantages. First, private data sharing is possible among different entities and second, this enables the training of complicated situations

Paper	Author	Date	Abstract	Application
Privacy-Preserving Traffic Flow Prediction: A Federated Learning Approach	Liu Y., Yu J.J.Q., Kang J., Niyato D., Zhang S.	2020	Existing traffic flow forecasting approaches by deep learning models achieve excellent success based on a large volume of data sets gathered by governments and organizations. However, these data sets may contain lots of user's private data, which is challenging the current prediction approaches as user privacy is calling for the public concern in recent years. Therefore, how to develop accurate traffic prediction while preserving privacy is a significant problem to be solved, and there is a tradeoff between these two objectives. To address this challenge, we introduce a privacy-preserving machine learning technique named federated learning (FL) and propose an FL-based gated recurrent unit neural network algorithm (FedGRU) for traffic flow prediction (TFP). FedGRU differs from current centralized learning methods and updates universal learning models through a secure parameter aggregation mechanism rather than directly sharing raw data among organizations. In the secure parameter aggregation mechanism, we adopt a federated averaging algorithm to reduce the communication overhead during the model parameter transmission process. Furthermore, we design a joint announcement protocol to improve the scalability of FedGRU. We also propose an ensemble clustering-based scheme for TFP by grouping the organizations into clusters before applying the FedGRU algorithm. Extensive case studies on a real-world data set demonstrate that FedGRU can produce predictions that are merely 0.76 km/h worse than the state of the art in terms of mean average error under the privacy preservation constraint, confirming that the proposed model develops accurate traffic predictions without compromising the data privacy. © 2014 IEEE.	Traffic flow prediction
Federated Learning with Blockchain for Autonomous Vehicles: Analysis and Design Challenges	Pokhrel S.R., Choi J.	2019	We propose an autonomous blockchain-based federated learning (BFL) design for privacy-aware and efficient vehicular communication networking, where local on-vehicle machine learning (oVML) model updates are exchanged and verified in a distributed fashion. BFL enables oVML without any centralized training data or coordination by utilizing the consensus mechanism of the blockchain. Relying on a renewal reward approach, we develop a mathematical framework that features the controllable network and BFL parameters (e.g., the retransmission limit, block size, block arrival rate, and the frame sizes) so as to capture their impact on the system-level performance. More importantly, our rigorous analysis of oVML system dynamics quantifies the end-to-end delay with BFL, which provides important insights into deriving optimal block arrival rate by considering communication and consensus delays. We present a variety of numerical and simulation results highlighting various non-trivial findings and insights for adaptive BFL design. In particular, based on analytical results, we minimize the system delay by exploiting the channel dynamics and demonstrate that the proposed idea of tuning the block arrival rate is provably online and capable of driving the system dynamics to the desired operating point. It also identifies the improved dependency on other blockchain parameters for a given set of channel conditions, retransmission limits, and frame sizes. However, a number of challenges (gaps in knowledge) need to be resolved in order to realise these changes. In particular, we identify key bottleneck challenges requiring further investigations, and provide potential future reserach directions. An early version of this work has been accepted for presentation in IEEE WCNC Wksps 2020 [1]. © 2020 IEEE.	Vehicle to vehicle communication

Paper	Author	Date	Abstract	Application
Deep-Reinforcement-Learning-Based Mode Selection and Resource Allocation for Cellular V2X Communications	Zhang X., Peng M., Yan S., Sun Y.	2020	Cellular vehicle-to-everything (V2X) communication is crucial to support future diverse vehicular applications. However, for safety-critical applications, unstable vehicle-to-vehicle (V2V) links, and high signaling overhead of centralized resource allocation approaches become bottlenecks. In this article, we investigate a joint optimization problem of transmission mode selection and resource allocation for cellular V2X communications. In particular, the problem is formulated as a Markov decision process, and a deep reinforcement learning (DRL)-based decentralized algorithm is proposed to maximize the sum capacity of vehicle-to-infrastructure users while meeting the latency and reliability requirements of V2V pairs. Moreover, considering training limitation of local DRL models, a two-timescale federated DRL algorithm is developed to help obtain robust models. Wherein, the graph theory-based vehicle clustering algorithm is executed on a large timescale and in turn, the federated learning algorithm is conducted on a small timescale. The simulation results show that the proposed DRL-based algorithm outperforms other decentralized baselines, and validate the superiority of the two-timescale federated DRL algorithm for newly activated V2V pairs. © 2014 IEEE.	Vehicle to vehicle communication
Joint Intelligence Ranking by Federated Multiplicative Update	Zhang C., Liu Y., Wang L., Liu Y., Li L., Zheng N.	2020	The joint intelligence ranking of intelligent systems like autonomous driving is of great importance for building a more general, extensive, and universally accepted intelligence evaluation scheme. However, due to issues such as privacy security and industry or area competition, the integration of isolated test results may face large unimaginable difficulty in information security and encrypted model training. To address this, we derive the federated multiplicative update (FMU) algorithm with boundary constraints to solve the nonnegative matrix factorization based joint intelligence ranking. The encrypted learning process is developed to alternate original computation steps in multiplicative update algorithms. Owning feasible property for the fast convergence and secure exchange of variables, the proposed framework outperforms the previous work on both real and simulated data. Further experimental analysis reveals that the introduced federated mechanism does not harm the overall time efficiency. © 2001-2011 IEEE.	Intelligence Ranking
An improved traffic congestion monitoring system based on federated learning	Xu C., Mao Y.	2020	This study introduces a software-based traffic congestionmonitoring system. The transportation system controls the traffic between cities all over the world. Traffic congestion happens not only in cities, but also on highways and other places. The current transportation system is not satisfactory in the area without monitoring. In order to improve the limitations of the current traffic system in obtaining road data and expand its visual range, the system uses remote sensing data as the data source for judging congestion. Since some remote sensing data needs to be kept confidential, this is a problem to be solved to effectively protect the safety of remote sensing data during the deep learning training process. Compared with the general deep learning training method, this study provides a federated learning method to identify vehicle targets in remote sensing images to solve the problem of data privacy in the training process of remote sensing data. The experiment takes the remote sensing image data sets of Los Angeles Road and Washington Road as samples for training, and the training results can achieve an accuracy of about 85%, and the estimated processing time of each image can be as low as 0.047 s. In the final experimental results, the system can automatically identify the vehicle targets in the remote sensing images to achieve the purpose of detecting congestion. © 2020 by the authors.	Traffic flow prediction

Paper	Author	Date	Abstract	Application
Inter-stakeholders Relationship in the Envisioned Autonomous Driving Era	Khan M.A., Kulkarni P., El Sayed H.	2020	Autonomous vehicles are expected to arrive sooner than expected. Autonomous vehicles of higher automation rely on both on-board and on-road deployed sensory data. Advanced approaches to improve the situational awareness of autonomous vehicle suggest to implement federated learning, where the raw data need to be transferred from vehicles to edges or clouds and vice-versa. This consequently generates dynamically varying communication link demands. The envisioned new era of autonomous driving demands the strong interplay of key stakeholders like: city authorities and communication network providers. In this paper, we study this relationship, where the traffic efficiency on different road segments may be achieved by incentivizing the autonomous vehicles through better communication resources on alternate routes. We model profit functions of the involved stakeholder. To carryout experiments, we use real traffic data of 8 months, which were collected through sensors deployed at Ernst-Reuter-Platz, Berlin, Germany. We developed an extensive validation framework to validate the approach, which comprises of SUMO, network simulator, and contributed modules. Results show that proposed approach achieves the traffic efficiency and help network operators to use the under-utilized network resources on the alternate paths. © 2020 ACM.	Traffic flow prediction
Improving TCP Performance over WiFi for Internet of Vehicles: A Federated Learning Approach	Pokhrel S.R., Choi J.	2020	We propose a novel communication efficient and privacy preserving federated learning framework for enhancing the performance of Internet of Vehicles (IoV), wherein on-vehicle learning models are trained by exchanging inputs, outputs and their learning parameters locally. Moreover, we use analytic modeling as a tool for reasoning and developing the required IoV scenario and stabilize their data flow dynamics by considering TCP CUBIC streams over WiFi networks to prove our idea. © 1967-2012 IEEE.	Vehicle to vehicle communication
Exploiting Unlabeled Data in Smart Cities using Federated Edge Learning	Albaseer A., Ciftler B.S., Abdallah M., Al-Fuqaha A.	2020	Privacy concerns are considered one of the main challenges in smart cities as sharing sensitive data induces threatening problems in people's lives. Federated learning has emerged as an effective technique to avoid privacy infringement as well as increase the utilization of the data. However, there is a scarcity in the amount of labeled data and an abundance of unlabeled data collected in smart cities; hence there is a necessity to utilize semi-supervised learning. In this paper, we present the primary design aspects for enabling federated learning at the edge networks taking into account the problem of unlabeled data. We propose a semi-supervised federated edge learning method called FedSem that exploits unlabeled data in real-time. FedSem algorithm is divided into two phases. The first phase trains a global model using only the labeled data. In the second phase, Fedsem injects unlabeled data into the learning process using the pseudo labeling technique and the model developed in the first phase to improve the learning performance. We carried out several experiments using the traffic signs dataset as a case study. Our results show that FedSem can achieve accuracy by up to 8% by utilizing the unlabeled data in the learning process. © 2020 IEEE.	Traffic flow prediction
Learning Cooperation Schemes for Mobile Edge Computing Empowered Internet of Vehicles	Cao J., Zhang K., Wu F., Leng S.	2020	Intelligent Transportation System has emerged as a promising paradigm providing efficient traffic management while enabling innovative transport services. The implementation of ITS always demands intensive computation processing under strict delay constraints. Machine Learning empowered Mobile Edge Computing (MEC), which brings intelligent computing service to the proximity of smart vehicles, is a potential approach to meet the processing demands. However, directly offloading and calculating these computation tasks in MEC servers may seriously impair the privacy of end users. To address this problem, we leverage federated learning in MEC empowered internet of vehicles to protect task data privacy. Moreover, we propose optimized learning cooperation schemes, which adaptively take smart vehicles and road side units to act as learning agents, and significantly reduce the learning costs in task execution. Numerical results demonstrate the effectiveness of our schemes. © 2020 IEEE.	Vehicle to vehicle communication

Paper	Author	Date	Abstract	Application
Time-dependent decentralized routing using federated learning	Wilbur M., Samal C., Talusan J.P., Yasumoto K., Dubey A.	2020	Recent advancements in cloud computing have driven rapid development in data-intensive smart city applications by providing near real time processing and storage scalability. This has resulted in efficient centralized route planning services such as Google Maps, upon which millions of users rely. Route planning algorithms have progressed in line with the cloud environments in which they run. Current state of the art solutions assume a shared memory model, hence deployment is limited to multiprocessing environments in data centers. By centralizing these services, latency has become the limiting parameter in the technologies of the future, such as autonomous cars. Additionally, these services require access to outside networks, raising availability concerns in disaster scenarios. Therefore, this paper provides a decentralized route planning approach for private fog networks. We leverage recent advances in federated learning to collaboratively learn shared prediction models online and investigate our approach with a simulated case study from a mid-size U.S. city. © 2020 IEEE.	Navigation
Federated Learning for Data Privacy Preservation in Vehicular Cyber-Physical Systems	Lu Y., Huang X., Dai Y., Maharjan S., Zhang Y.	2020	Recent developments in technologies such as MEC and AI contribute significantly in accelerating the deployment of VCPS. Techniques such as dynamic content caching, efficient resource allocation, and data sharing play a crucial role in enhancing the service quality and user driving experience. Meanwhile, data leakage in VCPS can lead to physical consequences such as endangering passenger safety and privacy, and causing severe property loss for data providers. The increasing volume of data, the dynamic network topology, and the availability of limited resources make data leakage in VCPS an even more challenging problem, especially when it involves multiple users and multiple transmission channels. In this article, we first propose a secure and intelligent architecture for enhancing data privacy. Then we present our new privacy-preserving federated learning mechanism and design a two-phase mitigating scheme consisting of intelligent data transformation and collaborative data leakage detection. Numerical results based on a real-world dataset demonstrate the effectiveness of our proposed scheme and show that our scheme achieves good accuracy, efficiency, and high security. © 1986-2012 IEEE.	Vehicle to vehicle communication
A Decentralized Federated Learning Approach for Connected Autonomous Vehicles	Pokhrel S.R., Choi J.	2020	In this paper, we propose an autonomous blockchain-based federated learning (BFL) design for privacy-aware and efficient vehicular communication networking, where local on-vehicle machine learning (oVML) model updates are exchanged and verified in a distributed fashion. BFL enables on-vehicle machine learning without any centralized training data or coordination by utilizing the consensus mechanism of the blockchain. Relying on a renewal reward approach, we develop a mathematical framework that features the controllable network and BFL parameters, such as the retransmission limit, block size, block arrival rate, and the frame sizes, so as to capture their impact on the system-level performance. More importantly, our rigorous analysis of oVML system dynamics quantifies the end-to-end delay with BFL, which provides important insights into deriving optimal block arrival rate by considering communication and consensus delays. © 2020 IEEE.	Vehicle to vehicle communication
Blockchain Empowered Asynchronous Federated Learning for Secure Data Sharing in Internet of Vehicles	Lu Y., Huang X., Zhang K., Maharjan S., Zhang Y.	2020	In Internet of Vehicles (IoV), data sharing among vehicles for collaborative analysis can improve the driving experience and service quality. However, the bandwidth, security and privacy issues hinder data providers from participating in the data sharing process. In addition, due to the intermittent and unreliable communications in IoV, the reliability and efficiency of data sharing need to be further enhanced. In this paper, we propose a new architecture based on federated learning to relieve transmission load and address privacy concerns of providers. To enhance the security and reliability of model parameters, we develop a hybrid blockchain architecture which consists of the permissioned blockchain and the local Directed Acyclic Graph (DAG). Moreover, we propose an asynchronous federated learning scheme by adopting Deep Reinforcement Learning (DRL) for node selection to improve the efficiency. The reliability of shared data is also guaranteed by integrating learned models into blockchain and executing a two-stage verification. Numerical results show that the proposed data sharing scheme provides both higher learning accuracy and faster convergence. © 1967-2012 IEEE.	Internet of Vehicles

Paper	Author	Date	Abstract	Application
Poisoning Attacks in Federated Learning: An Evaluation on Traffic Sign Classification	Nuding F., Mayer R.	2020	Federated Learning has recently gained attraction as a means to analyze data without having to centralize it from initially distributed data sources. Generally, this is achieved by only exchanging and aggregating the parameters of the locally learned models. This enables better handling of sensitive data, e.g. of individuals, or business related content. Applications can further benefit from the distributed nature of the learning by using multiple computer resources, and eliminating network communication overhead. Adversarial Machine Learning in general deals with attacks on the learning process, and backdoor attacks are one specific attack that tries to break the integrity of a model by manipulating the behavior on certain inputs. Recent work has shown that despite the benefits of Federated Learning, the distributed setting also opens up new attack vectors for adversaries. In this paper, we thus specifically study this manipulation of the training process to embed a backdoor on the example of a dataset for traffic sign classification. Extending earlier work, we specifically include the setting of sequential learning, in additional to parallel averaging, and perform a broad analysis on a number of different settings. © 2020 ACM.	Traffic sign classification
Differentially private asynchronous federated learning for mobile edge computing in urban informatics	Lu Y., Huang X., Dai Y., Maharjan S., Zhang Y.	2019	Driven by technologies such as mobile edge computing and 5G, recent years have witnessed the rapid development of urban informatics, where a large amount of data is generated. To cope with the growing data, artificial intelligence algorithms have been widely exploited. Federated learning is a promising paradigm for distributed edge computing, which enables edge nodes to train models locally without transmitting their data to a server. However, the security and privacy concerns of federated learning hinder its wide deployment in urban applications such as vehicular networks. In this article, we propose a differentially private asynchronous federated learning scheme for resource sharing in vehicular networks. To build a secure and robust federated learning scheme, we incorporate local differential privacy into federated learning for protecting the privacy of updated local models. We further propose a random distributed update scheme to get rid of the security threats led by a centralized curator. Moreover, we perform the convergence boosting in our proposed scheme by updates verification and weighted aggregation. We evaluate our scheme on three real-world datasets. Numerical results show the high accuracy and efficiency of our proposed scheme, whereas preserve the data privacy. © 2005-2012 IEEE.	Vehicle to vehicle communication
Distributed Federated Learning for Ultra-Reliable Low-Latency Vehicular Communications	Samarakoon S., Bennis M., Saad W., Debbah M.	2019	In this paper, the problem of joint power and resource allocation (JPRA) for ultra-reliable low-latency communication (URLLC) in vehicular networks is studied. Therein, the network-wide power consumption of vehicular users (VUEs) is minimized subject to high reliability in terms of probabilistic queuing delays. Using extreme value theory (EVT), a new reliability measure is defined to characterize extreme events pertaining to vehicles' queue lengths exceeding a predefined threshold. To learn these extreme events, assuming they are independently and identically distributed over VUEs, a novel distributed approach based on federated learning (FL) is proposed to estimate the tail distribution of the queue lengths. Considering the communication delays incurred by FL over wireless links, Lyapunov optimization is used to derive the JPRA policies enabling URLLC for each VUE in a distributed manner. The proposed solution is then validated via extensive simulations using a Manhattan mobility model. Simulation results show that FL enables the proposed method to estimate the tail distribution of queues with an accuracy that is close to a centralized solution with up to 79% reductions in the amount of exchanged data. Furthermore, the proposed method yields up to 60% reductions of VUEs with large queue lengths, while reducing the average power consumption by two folds, compared to an average queue-based baseline. © 2019 IEEE.	Vehicle to vehicle communication

Paper	Author	Date	Abstract	Application
Federated Learning for Vehicular Internet of Things: Recent Advances and Open Issues	Du Z., Wu C., Yoshinaga T., Yau K.A., Ji Y., Li J.	2020	Federated learning (FL) is a distributed machine learning approach that can achieve the purpose of collaborative learning from a large amount of data that belong to different parties without sharing the raw data among the data owners. FL can sufficiently utilize the computing capabilities of multiple learning agents to improve the learning efficiency while providing a better privacy solution for the data owners. FL attracts tremendous interests from a large number of industries due to growing privacy concerns. Future vehicular Internet of Things (IoT) systems, such as cooperative autonomous driving and intelligent transport systems (ITS), feature a large number of devices and privacy-sensitive data where the communication, computing, and storage resources must be efficiently utilized. FL could be a promising approach to solve these existing challenges. In this paper, we first conduct a brief survey of existing studies on FL and its use in wireless IoT. Then we discuss the significance and technical challenges of applying FL in vehicular IoT, and point out future research directions. CCBY	Vehicle to vehicle communication
Energy demand prediction with federated learning for electric vehicle networks	Saputra Y.M., Hoang D.T., Nguyen D.N., Dutkiewicz E., Mueck M.D., Srikanteswara S.	2019	In this paper, we propose novel approaches using state-of-the-art machine learning techniques, aiming at predicting energy demand for electric vehicle (EV) networks. These methods can learn and find the correlation of complex hidden features to improve the prediction accuracy. First, we propose an energy demand learning (EDL)-based prediction solution in which a charging station provider (CSP) gathers information from all charging stations (CSs) and then performs the EDL algorithm to predict the energy demand for the considered area. However, this approach requires frequent data sharing between the CSs and the CSP, thereby driving communication overhead and privacy issues for the EVs and CSs. To address this problem, we propose a federated energy demand learning (FEDL) approach which allows the CSs sharing their information without revealing real datasets. Specifically, the CSs only need to send their trained models to the CSP for processing. In this case, we can significantly reduce the communication overhead and effectively protect data privacy for the EV users. To further improve the effectiveness of the FEDL, we then introduce a novel clustering- based EDL approach for EV networks by grouping the CSs into clusters before applying the EDL algorithms. Through experimental results, we show that our proposed approaches can improve the accuracy of energy demand prediction up to 24.63% and decrease communication overhead by 83.4% compared with other baseline machine learning algorithms. © 2019 IEEE.	Energy demand prediction
Federated Learning for Ultra-Reliable Low-Latency V2V Communications	Samarakoon S., Bennis M., Saad W., Debbah M.	2018	In this paper, a novel joint transmit power and resource allocation approach for enabling ultra-reliable low-latency communication (URLLC) in vehicular networks is proposed. The objective is to minimize the network-wide power consumption of vehicular users (VUEs) while ensuring high reliability in terms of probabilistic queuing delays. In particular, a reliability measure is defined to characterize extreme events (i.e., when vehicles' queue lengths exceed a predefined threshold with non-negligible probability) using extreme value theory (EVT). Leveraging principles from federated learning (FL), the distribution of these extreme events corresponding to the tail distribution of queues is estimated by VUEs in a decentralized manner. Finally, Lyapunov optimization is used to find the joint transmit power and resource allocation policies for each VUE in a distributed manner. The proposed solution is validated via extensive simulations using a Manhattan mobility model. It is shown that FL enables the proposed distributed method to estimate the tail distribution of queues with an accuracy that is very close to a centralized solution with up to 79% reductions in the amount of data that need to be exchanged. Furthermore, the proposed method yields up to 60% reductions of VUEs with large queue lengths, without an additional power consumption, compared to an average queue-based baseline. Compared to systems with fixed power consumption and focusing on queue stability while minimizing average power consumption, the reductions in extreme events of the proposed method is about two orders of magnitude. © 2018 IEEE.	Vehicle to vehicle communication

Paper	Author	Date	Abstract	Application
Federated Learning in Vehicular Networks	Ahmet M. Elbir, Burak Soner, Sinem Coleri	2020	Machine learning (ML) has already been adopted in vehicular networks for such applications as autonomous driving, road safety prediction and vehicular object detection, due to its model-free characteristic, allowing adaptive fast response. However, the training of the ML model brings significant overhead for the data transmission between the parameter server and the edge devices in the vehicles. Federated learning (FL) framework has been recently introduced as an efficient tool with the goal of reducing this transmission overhead while also achieving privacy through the transmission of only the model updates of the learnable parameters rather than the whole dataset. In this article, we investigate the usage of FL over ML in vehicular network applications to develop intelligent transportation systems. We provide a comprehensive analysis on the feasibility of FL for the ML based vehicular applications. Then, we identify the major challenges from both learning perspective, i.e., data labeling and model training, and from the communications point of view, i.e., data rate, reliability, transmission overhead/delay, privacy and resource management. Finally, we highlight related future research directions for FL in vehicular networks. © arXiv:2006.01412 [eess.SP]	Intelligent transportation systems
Federated Learning in Vehicular Edge Computing: A Selective Model Aggregation Approach	D. Ye, R. Yu, M. Pan and Z. Han	21 January 2020	Federated learning is a newly emerged distributed machine learning paradigm, where the clients are allowed to individually train local deep neural network (DNN) models with local data and then jointly aggregate a global DNN model at the central server. Vehicular edge computing (VEC) aims at exploiting the computation and communication resources at the edge of vehicular networks. Federated learning in VEC is promising to meet the ever-increasing demands of artificial intelligence (AI) applications in intelligent connected vehicles (ICV). Considering image classification as a typical AI application in VEC, the diversity of image quality and computation capability in vehicular clients potentially affects the accuracy and efficiency of federated learning. Accordingly, we propose a selective model aggregation approach, where "fine" local DNN models are selected and sent to the central server by evaluating the local image quality and computation capability. Regarding the implementation of model selection, the central server is not aware of the image quality and computation capability in the vehicular clients, whose privacy is protected under such a federated learning framework. To overcome this information asymmetry, we employ two-dimension contract theory as a distributed framework to facilitate the interactions between the central server and vehicular clients. The formulated problem is then transformed into a tractable problem through successively relaxing and simplifying the constraints, and eventually solved by a greedy algorithm. Using two datasets, i.e., MNIST and BelgiumTSC, our selective model aggregation approach is demonstrated to outperform the original federated averaging (FedAvg) approach in terms of accuracy and efficiency. Meanwhile, our approach also achieves higher utility at the central server compared with the baseline approaches. © 2020 IEEE.	Centralized data exchange

Paper	Author	Date	Abstract	Application
Blockchain-Supported Federated Learning for Trustworthy Vehicular Networks	S. Otoum, I. Al Ridhawi and H. T. Mouftah	2020	The advances in today's IoT devices and machine learning methods have given rise to the concept of Federated Learning. Through such a technique, a plethora of network devices collaboratively train and update a mutual machine learning model while protecting their individual datasets. Federated learning proves its effectiveness in tackling communication efficiency and privacy-safeguarding issues. Moreover, blockchain was introduced to solve many network issues in regard to data privacy and network single point of failure. In this article, we introduce a solution that integrates both federated learning and blockchain to ensure both data privacy and network security. We present a framework to decentralize the mutual machine learning models on end-devices. A blockchain-based consensus solution as a second line of privacy is used to ensure trustworthy shared training on the fog. The proposed model enables on-end device machine learning without any centralized training of the data nor coordination by utilizing a consensus method in the blockchain. We evaluate and verify our proposed model through simulation to showcase the effectiveness of the adapted scheme in terms of accuracy, energy consumption, and lifetime rate, along with throughput and latency metrics. The proposed model performs with an accuracy rate of ≈ 0.97. © 2020 IEEE	Trustworthy vehicular network
Federated Learning in Vehicular Networks: Opportunities and Solutions	J. Posner, L. Tseng, M. Aloqaily and Y. Jararweh	22 February 2021	The emerging advances in personal devices and privacy concerns have given the rise to the concept of Federated Learning. Federated Learning proves its effectiveness and privacy preservation through collaborative local training and updating a shared machine learning model while protecting the individual data-sets. This article investigates a new type of vehicular network concept, namely a Federated Vehicular Network (FVN), which can be viewed as a robust distributed vehicular network. Compared to traditional vehicular networks, an FVN has centralized components and utilizes both DSRC and mmWave communication to achieve more scalable and stable performance. As a result, FVN can be used to support data-/computation-intensive applications such as distributed machine learning and Federated Learning. The article first outlines the enabling technologies of FVN. Then, we briefly discuss the high-level architecture of FVN and explain why such an architecture is adequate for Federated Learning. In addition, we use auxiliary Blockchain-based systems to facilitate transactions and mitigate malicious behaviors. Next, we discuss in detail one key component of FVN, a federated vehicular cloud (FVC), that is used for sharing data and models in FVN. In particular, we focus on the routing inside FVCs and present our solutions and preliminary evaluation results. Finally, we point out open problems and future research directions of this disruptive technology. ©2021 IEEE	Vehicular network concept
Energy-Aware Blockchain and Federated Learning-Supported Vehicular Networks	M. Aloqaily, I. A. Ridhawi and M. Guizani	17. Aug 21	The aerial capabilities and flexibility in movement of Unmanned Aerial Vehicles (UAVs) has enabled them to adaptively provide both traditional and more contemporary services. In this article, we introduce a solution that integrates the capabilities of both UAVs and Unmanned Ground Vehicles (UGVs) to provide both intelligent connectivity and services to both aerial and ground connected devices. A cooperative solution is adopted that considers nodes' power and movement constraints. The UAV and UGV cooperative process ensures continuous power availability to UAVs to support seamless and continuous service availability to end-devices. A Federated Learning (FL) approach is adopted at the edge to ensure accurate and up-to-date service provisioning in accordance with the surrounding environment and network constraints. Moreover, Blockchain technology is used to decentralize the provisioning and control aspects, and ensure authenticity and integrity. Extensive simulations are conducted to test the soundness and applicability of the proposed solution. Results show significant improvement in terms of connectivity, service availability, and UAV energy enhancements when compared to traditional mobile and vehicular communication techniques. ©2021 IEEE	Vehicular communication

Paper	Author	Date	Abstract	Application
Federated Learning Meets Contract Theory: Economic-Efficiency Framework for Electric Vehicle Networks	Y. M. Saputra, D. Nguyen, H. T. Dinh, T. X. Vu, E. Dutkiewicz and S. Chatzinotas	21 December 2020	In this paper, we propose a novel economic-efficiency framework for an electric vehicle (EV) network to maximize the profits for charging stations (CSs). To that end, we first introduce an energy demand prediction method for CSs leveraging federated learning approaches, in which each CS can exchange its learned model with other CSs to improve the learning quality. Based on the predicted energy demands, each CS can reserve energy from the smart grid provider (SGP) in advance to optimize its profit. Nonetheless, due to the competition among the CSs as well as unknown information from the SGP, we develop a multi-principal one-agent (MPOA) contract-based method to address these issues. In particular, we formulate the CSs' profit maximization as a non-collaborative energy contract problem under the SGP's unknown information and common constraints, and other CSs' contracts. To solve this problem, we transform it into an equivalent low-complexity problem and develop an iterative algorithm to find the optimal contracts for the CSs. Through simulation results, we demonstrate that our proposed framework can enhance energy demand prediction accuracy up to 24.63% compared with other machine learning algorithms. Furthermore, our proposed framework can outperform other economic models by 48% and 36% in terms of the CSs' utilities and social welfare of the network, respectively. ©2020 IEEE	Communication among charging stations
Federated Learning Meets Contract Theory: Energy-Efficient Framework for Electric Vehicle Networks	Yuris Mulya Saputra, Diep N. Nguyen, Dinh Thai Hoang, Thang Xuan Vu, Eryk Dutkiewicz, Symeon Chatzinotas	2020	In this paper, we propose a novel energy-efficient framework for an electric vehicle (EV) network using a contract theoretic-based economic model to maximize the profits of charging stations (CSs) and improve the social welfare of the network. Specifically, we first introduce CS-based and CS clustering-based decentralized federated energy learning (DFEL) approaches which enable the CSs to train their own energy transactions locally to predict energy demands. In this way, each CS can exchange its learned model with other CSs to improve prediction accuracy without revealing actual datasets and reduce communication overhead among the CSs. Based on the energy demand prediction, we then design a multi-principal one-agent (MPOA) contract-based method. In particular, we formulate the CSs' utility maximization as a non-collaborative energy contract problem in which each CS maximizes its utility under common constraints from the smart grid provider (SGP) and other CSs' contracts. Then, we prove the existence of an equilibrium contract solution for all the CSs and develop an iterative algorithm at the SGP to find the equilibrium. Through simulation results using the dataset of CSs' transactions in Dundee city, the United Kingdom between 2017 and 2018, we demonstrate that our proposed method can achieve the energy demand prediction accuracy improvement up to 24.63% and lessen communication overhead by 96.3% compared with other machine learning algorithms. Furthermore, our proposed method can outperform non-contract-based economic models by 35% and 36% in terms of the CSs' utilities and social welfare of the network, respectively. ©arXiv:2004.01828	Decentralized communication among charging stations
Secure-Enhanced Federated Learning for AI-Empowered Electric Vehicle Energy Prediction	W. Wang, M. H. Fida, Z. Lian, Z. Yin, Q-V. Pham, T. R. Gadekallu, K. Dev, C. Su.	30. Sep 21	Although AI-empowered schemes bring some sound solutions to stimulate more reasonable energy distribution schemes between charging stations (CSs) and a charging stationproviders (CSP), frequent data sharing between them is possible to incur many security and privacy breaches. To solve these problems, federated learning (FL) is an ideal solution thatonly requires CSs to upload local models instead of detailed data. Although the CSs electricity consumption needs not to beexposed to the server directly, FL-based schemes still have been excavated several security threats such as information exploiting attacks, data poisoning attacks, model poisoning attacks, andfree-riding attacks. Hence, in this paper, both the effectiveness of energy management and the potential risks of FL for electricvehicle infrastructures (EVIs) are considered, we propose alightweight authentication FL-based energy demand predictionfor EVIs with premium-penalty mechanism. Security analysis andperformance evaluation prove that our proposed framework cangenerate an accurate electricity demand prediction framework to defend multiple FL attacks for EVIs.© 2021 IEEE Consumer Electronics Magazine	Energy demand prediction

Paper	Author	Date	Abstract	Application
Privacy-preserving blockchain-based federated learning for traffic flow prediction	Yuanhang Qi, M. Shamim Hossain, Jiangtian Nie, Xuandi Li	2020	As accurate and timely traffic flow information is extremely important for traffic management, traffic flow prediction has become a vital component of intelligent transportation systems. However, existing traffic flow prediction methods based on centralized machine learning need to gather raw data for model training, which involves serious privacy exposure risks. To address these problems, federated learning that shares model updates without exchanging raw data, has recently been introduced as an efficient solution for achieving privacy protection. However, the existing federated learning frameworks are based on a centralized model coordinator that still suffers from severe security challenges, such as a single point of failure. Thereby, a consortium blockchain-based federated learning framework is proposed to enable decentralized, reliable, and secure federated learning without a centralized model coordinator. In the proposed framework, the model updates from distributed vehicles are verified by miners to prevent unreliable model updates and are then stored on the blockchain. In addition, to further protect model privacy on the blockchain, a differential privacy method with a noise-adding mechanism is applied for the blockchain-based federated learning framework. Numerical results illustrate that the proposed schemes can effectively prevent data poisoning attacks and improve the privacy protection of model updates for secure and privacy-preserving traffic flow prediction. © 2020 Elsevier B.V.	Traffic flow prediction
Dual Attention-Based Federated Learning for Wireless Traffic Prediction	C. Zhang, S. Dang, B. Shihada and M. -S. Alouini	2021	Wireless traffic prediction is essential for cellular networks to realize intelligent network operations, such as load-aware resource management and predictive control. Existing prediction approaches usually adopt centralized training architectures and require the transferring of huge amounts of traffic data, which may raise delay and privacy concerns for certain scenarios. In this work, we propose a novel wireless traffic prediction framework named Dual Attention-Based Federated Learning (FedDA), by which a high-quality prediction model is trained collaboratively by multiple edge clients. To simultaneously capture the various wireless traffic patterns and keep raw data locally, FedDA first groups the clients into different clusters by using a small augmentation dataset. Then, a quasi-global model is trained and shared among clients as prior knowledge, aiming to solve the statistical heterogeneity challenge confronted with federated learning. To construct the global model, a dual attention scheme is further proposed by aggregating the intra-and inter-cluster models, instead of simply averaging the weights of local models. We conduct extensive experiments on two real-world wireless traffic datasets and results show that FedDA outperforms state-of-the-art methods. The average mean squared error performance gains on the two datasets are up to 10% and 30%, respectively. ©2021 IEEE	Traffic prediction
Multi-Task Federated Learning for Traffic Prediction and Its Application to Route Planning	T. Zeng; J. Guo; K. J. Kim; K. Parsons; P. Orlik; S. Di Cairano;W. Saad.	2021	A novel multi-task federated learning (FL) framework is proposed in this paper to optimize the traffic prediction models without sharing the collected data among traffic stations. In particular, a divisive hierarchical clustering is first introduced to partition the collected traffic data at each station into different clusters. The FL is then implemented to collaboratively train the learning model for each cluster of local data distributed across the stations. Using the multi-task FL framework, the route planning is studied where the road map is modeled as a time-dependent graph and a modified A * algorithm is used to determine the route with the shortest traveling time. Simulation results showcase the prediction accuracy improvement of the proposed multi-task FL framework over two baseline schemes. The simulation results also show that, when using the multi-task FL framework in the route planning, an accurate traveling time can be estimated and an effective route can be selected.©2021 IEEE	Traffic prediction

Paper	Author	Date	Abstract	Application
A Federated Learning Approach to Routing in Challenged SDN-Enabled Edge Networks	A. Sacco, F. Esposito and G. Marchetto	2020	The edge computing paradigm allows computationally intensive tasks to be offloaded from small devices to nearby (more) powerful servers, via an edge network. The intersection between such edge computing paradigm and Machine Learning (ML), in general, and deep learning in particular, has brought to light several advantages for network operators: from automating management tasks, to gain additional insights on their networks. Most of the existing approaches that use ML to drive routing and traffic control decisions are valuable but rarely focus on challenged networks, that are characterized by continually varying network conditions and the high volume of traffic generated by edge devices. In particular, recently proposed distributed ML-based architectures require either a long synchronization phase or a training phase that is unsustainable for challenged networks. In this paper, we fill this knowledge gap with Blaster, a federated architecture for routing packets within a distributed edge network, to improve the application's performance and allow scalability of data-intensive applications. We also propose a novel path selection model that uses Long Short Term Memory (LSTM) to predict the optimal route. Finally, we present some initial results obtained by testing our approach via simulations and with a prototype deployed over the GENI testbed. By leveraging a Federated Learning (FL) model, our approach shows that we can optimize the communication between SDN controllers, preserving bandwidth for the data traffic.©2020 IEEE	Traffic prediction, optimal route
FASTGNN: A Topological Information Protected Federated Learning Approach for Traffic Speed Forecasting	C. Zhang, S. Zhang, J. J. Q. Yu and S. Yu	Dec.2021	Federated learning has been applied to various tasks in intelligent transportation systems to protect data privacy through decentralized training schemes. The majority of the state-of-the-art models in intelligent transportation systems (ITS) are graph neural networks (GNN)-based for spatial information learning. When applying federated learning to the ITS tasks with GNN-based models, the existing frameworks can only protect the data privacy; however, ignore the one of topological information of transportation networks. In this article, we propose a novel federated learning framework to tackle this problem. Specifically, we introduce a differential privacy-based adjacency matrix preserving approach for protecting the topological information. We also propose an adjacency matrix aggregation approach to allow local GNN-based models to access the global network for a better training effect. Furthermore, we propose a GNN-based model named attention-based spatial-temporal graph neural networks (ASTGNN) for traffic speed forecasting. We integrate the proposed federated learning framework and ASTGNN as FASTGNN for traffic speed forecasting. Extensive case studies on a real-world dataset demonstrate that FASTGNN can develop accurate forecasting under the privacy preservation constraint.©2021 IEEE	Traffic prediction
A Hierarchical Blockchain-Enabled Federated Learning Algorithm for Knowledge Sharing in Internet of Vehicles	H. Chai, S. Leng, Y. Chen and K. Zhang	July 2021	Internet of Vehicles (IoVs) is highly characterized by collaborative environment data sensing, computing and processing. Emerging Big Data and Artificial Intelligence (AI) technologies show significant advantages and efficiency for knowledge sharing among intelligent vehicles. However, it is challenging to guarantee the security and privacy of knowledge during the sharing process. Moreover, conventional AI-based algorithms cannot work properly in distributed vehicular networks. In this paper, a hierarchical blockchain framework and a hierarchical federated learning algorithm are proposed for knowledge sharing, by which vehicles learn environmental data through machine learning methods and share the learning knowledge with each others. The proposed hierarchical blockchain framework is feasible for the large scale vehicular networks. The hierarchical federated learning algorithm is designed to meet the distributed pattern and privacy requirement of IoVs. Knowledge sharing is then modeled as a trading market process to stimulate sharing behaviours, and the trading process is formulated as a multi-leader and multi-player game. Simulation results show that the proposed hierarchical algorithm can improve the sharing efficiency and learning quality. Furthermore, the blockchain-enabled framework is able to deal with certain malicious attacks effectively. ©2021 IEEE	Vehicular communication

Paper	Author	Date	Abstract	Application
Dynamic Federated Learning-Based Economic Framework for Internet-of-Vehicles	Yuris Mulya Saputra, Dinh Thai Hoang, Diep N. Nguyen, Le-Nam Tran, Shimin Gong, Eryk Dutkiewicz	2021	Federated learning (FL) can empower Internet-of-Vehicles (IoV) networks by leveraging smart vehicles (SVs) to participate in the learning process with minimum data exchanges and privacy disclosure. The collected data and learned knowledge can help the vehicular service provider (VSP) improve the global model accuracy, e.g., for road safety as well as better profits for both VSP and participating SVs. Nonetheless, there exist major challenges when implementing the FL in IoV networks, such as dynamic activities and diverse quality-of-information (QoI) from a large number of SVs, VSP's limited payment budget, and profit competition among SVs. In this paper, we propose a novel dynamic FL-based economic framework for an IoV network to address these challenges. Specifically, the VSP first implements an SV selection method to determine a set of the best SVs for the FL process according to the significance of their current locations and information history at each learning round. Then, each selected SV can collect on-road information and offer a payment contract to the VSP based on its collected QoI. For that, we develop a multi-principal one-agent contract-based policy to maximize the profits of the VSP and learning SVs under the VSP's limited payment budget and asymmetric information between the VSP and SVs. Through experimental results using real-world on-road datasets, we show that our framework can converge 57% faster (even with only 10% of active SVs in the network) and obtain much higher social welfare of the network (up to 27.2 times) compared with those of other baseline FL methods. ©arXiv:2101.00191	Vehicular communication
Joint resource management for mobility supported federated learning in Internet of Vehicles	Ge Wang, Fangmin Xu, Hengsheng Zhang, Chenglin Zhaoa	2021	In recent years, the powerful combination of Multi-access Edge Computing (MEC) and Artificial Intelligence (AI), called edge intelligence, promotes the development of Intelligent Transportation Systems (ITS). However, there is a mismatch between the ever-increasing consumer privacy awareness and the data leakage risk in centralized AI training solutions in vehicular edge scenarios, which has become a new obstacle to satisfying the user experience. As a promising privacy-preserving paradigm, federated learning synthesizes a global model only with the parameters of decentralized trained local models, avoiding the exposure of sensitive data. Given this, we introduce federated learning into the proposed two-level MEC-assisted vehicular network framework. This paper aims to address the challenges posed by adopting federated learning into the Internet of Vehicles (IoV) scenario. Firstly, as the entity of the participant (the local model training node of federated learning), vehicles have high mobility. We design a mobility supported federated learning participant decision algorithm to pick out participants from candidate vehicles. Secondly, federated learning is rather resource-consuming, inevitably incurring considerable costs to participants. We focus on the joint resource allocation problem to optimize the federated learning cost. Finally, considering the limitations of centralized resource allocation, we propose a fully distributed resource allocation method inspired by multi-agent deep reinforcement learning. Simulation results are presented to demonstrate the feasibility and effectiveness of the proposed schemes.© 2021 Published by Elsevier B.V.	Vehicular communication
BFLP: An Adaptive Federated Learning Framework for Internet of Vehicles	Y.Peng, Z. Chen, Z. Chen, W. Ou , W. Han, J. Ma	2021	Applications of Internet of Vehicles (IoV) make the life of human beings more intelligent and convenient. However, in the present, there are some problems in IoV, such as data silos and poor privacy preservation. To address the challenges in IoV, we propose a blockchain-based federated learning pool (BFLP) framework. BFLP allows the models to be trained without sharing raw data, and it can choose the most suitable federated learning method according to actual application scenarios. Considering the poor computing power of vehicle systems, we construct a lightweight encryption algorithm called CPC to protect privacy. To verify the proposed framework, we conducted experiments in obstacle-avoiding and traffic forecast scenarios. The results show that the proposed framework can effectively protect the user's privacy, and it is more stable and efficient compared with traditional machine learning technique. Also, we compare the CPC algorithm with other encryption algorithms. And the results show that its calculation cost is much lower compared to other symmetric encryption algorithms.© 2021 Yongqiang Peng et al.	Vehicular communication

Paper	Author	Date	Abstract	Application
A Federated Learning-Based License Plate Recognition Scheme for 5G-Enabled Internet of Vehicles	X.Kong, K.Wang, M.Hou,X.Hao, G.Shen, X.Chen, F.Xia	2021	License plate is an essential characteristic to identify vehicles for the traffic management, and thus, license plate recognition is important for Internet of Vehicles. Since 5G has been widely covered, mobile devices are utilized to assist the traffic management, which is a significant part of Industry 4.0. However, there have always been privacy risks due to centralized training of models. Also, the trained model cannot be directly deployed on the mobile device due to its large number of parameters. In this article, we propose a federated learning-based license plate recognition framework (FedLPR) to solve these problems. We design detection and recognition model to apply in the mobile device. In terms of user privacy, data in individuals is harnessed on their mobile devices instead of the server to train models based on federated learning. Extensive experiments demonstrate that FedLPR has high accuracy and acceptable communication cost while preserving user privacy. ©2021 IEEE	License plate recognition
FedVCP: A Federated-Learning-Based Cooperative Positioning Scheme for Social Internet of Vehicles	X. Kong, H. Gao, G. Shen, G. Duan and S. K. Das	2021	Intelligent vehicle applications, such as autonomous driving and collision avoidance, put forward a higher demand for precise positioning of vehicles. The current widely used global navigation satellite systems (GNSS) cannot meet the precision requirements of the submeter level. Due to the development of sensing techniques and vehicle-to-infrastructure (V2I) communications, some vehicles can interact with surrounding landmarks to achieve precise positioning. Existing work aims to realize the positioning correction of common vehicles by sharing the positioning data of sensor-rich vehicles. However, the privacy of trajectory data makes it difficult to collect and train data centrally. Moreover, uploading vehicle location data wastes network resources. To fill these gaps, this article proposes a vehicle cooperative positioning (CP) system based on federated learning (FedVCP), which makes full use of the potential of social Internet of Things (IoT) and collaborative edge computing (CEC) to provide high-precision positioning correction while ensuring user privacy. To the best of our knowledge, this article is the first attempt to solve the privacy of CP from a perspective of federated learning. In addition, we take the advantages of local cooperation through vehicle-to-vehicle (V2V) communications in data augmentation. For individual differences in vehicle positioning, we utilize transfer learning to eliminate the impact of such differences. Extensive experiments on real data demonstrate that our proposed model is superior to the baseline method in terms of effectiveness and convergence speed.©2021 IEEE	Vehicular communication
Decentralized Federated Learning for Road User Classification in Enhanced V2X Networks	L. Barbieri, S. Savazzi and M. Nicoli	2021	Federated Learning (FL) techniques are emerging in the automotive context to support connected automated driving services. Yet, when applied to vehicular use cases, conventional centralized FL policies show some drawbacks in terms of latency and scalability. This paper focuses on decentralized FL solutions, which attempt to overcome such limitations, by introducing a distributed computing architecture: vehicles exchange the parameters of a shared Machine Learning (ML) model via V2V links, without the need of a central orchestrator. Sharing all ML parameters, however, might not be feasible when minimal V2X bandwidth usage is required or the model is highly complex (e.g., extremely deep networks) as in advanced scenarios for high levels of automation. We thus propose a modular decentralized FL solution and we discuss its application to road user classification in a cooperative vehicular sensing use case. The proposed FL solution performs the point cloud processing of Lidar sensor inputs using a PointNet compliant architecture. It enables the exchange of a subset of the model parameters, namely selected ML model layers, optimized for communication efficiency, convergence and accuracy. We use real sensor data extracted from a publicly available dataset to validate the method, focusing on non-uniform scenarios where sensor data are highly unbalanced across the connected vehicles. For all cases, FL is shown to outperform the ego-sensing approach with minimal bandwidth usage.©2021 IEEE	Vehicular communication

Paper	Author	Date	Abstract	Application
Distributed Learning for Vehicle Routing Decision in Software Defined Internet of Vehicles	K. Lin, C. Li, Y. Li, C. Savaglio and G. Fortino	June 2021	With the increasing number of vehicles, the traffic congestion is becoming more and more serious. In order to alleviate such a problem, this article considers transmission and inference delay of cloud centralized computing in the software defined Internet of Vehicles (SDIoV), and builds a new SDIoV architecture based on edge intelligence, for supporting real-time vehicle routing decision through distributed multi-agent reinforcement learning model. Then, a software defined device collaboration optimization method is designed to improve the efficiency of distributed training. Combined with multi-agent reinforcement learning, a distributed-learning-based vehicle routing decision algorithm (DLRD) is proposed to adaptively adjust vehicle routing online. The performed simulations show that the DLRD can successfully realize real-time routing decision for vehicles and alleviate traffic congestion with the dynamic changes of the road environment.©2021 IEEE	Traffic prediction, optimal route
Joint Scheduling and Resource Allocation for Efficiency-Oriented Distributed Learning Over Vehicle Platooning Networks	X. Ma, J. Zhao and Y. Gong	Oct. 2021	The limited communication and computing resources, as well as the rising concerns about the privacy protection, bring significant challenges to the massive data training and analysis in vehicular networks. To address these challenges, in this paper a platoon-based distributed learning framework design for data learning is carried out, where the vacant computation resources of vehicle platooning networks are leveraged. In the proposed framework, a 2-phase Markovian stochastic process is utilized to depict the learning service heterogeneity for each participating vehicle. Meanwhile, we propose a joint scheduling and resource allocation scheme for efficiency-oriented distributed learning to maximize the learning accuracy subject to a given learning time constraint. The optimization problem is solved as follows. First, given the scheduled vehicles, the communication resource allocation is modeled as a minimum-maximum problem to minimize the learning delay of each learning round. Subsequently, an efficiency-oriented unbiased global aggregation policy is proposed to explore the convergence difference between partial scheduling and total scheduling. Considering the learning convergence and remaining time, an on-demand scheduling scheme is introduced to determine the number of scheduled vehicles. Finally, combining the learning efficiency of each vehicle with the scheduled number of vehicles, the scheduled vehicle set is selected. Simulations results show that the proposed scheduling policy can schedule the number of participating vehicles on demand based on the trade-off between learning performance and learning latency.©2021 IEEE	Vehicle platooning networks
FedParking: A Federated Learning Based Parking Space Estimation With Parked Vehicle Assisted Edge Computing	X. Huang, P. Li, R. Yu, Y. Wu, K. Xie and S. Xie	Sep 21	As a distributed learning approach, federated learning trains a shared learning model over distributed datasets while preserving the training data privacy. We extend the application of federated learning to parking management and introduce FedParking in which Parking Lot Operators (PLOs) collaborate to train a long short-term memory model for parking space estimation without exchanging the raw data. Furthermore, we investigate the management of Parked Vehicle assisted Edge Computing (PVEC) by FedParking. In PVEC, different PLOs recruit PVs as edge computing nodes for offloading services through an incentive mechanism, which is designed according to the computation demand and parking capacity constraints derived from FedParking. We formulate the interactions among the PLOs and vehicles as a multi-lead multi-follower Stackelberg game. Considering the dynamic arrivals of the vehicles and time-varying parking capacity constraints, we present a multi-agent deep reinforcement learning approach to gradually reach the Stackelberg equilibrium in a distributed yet privacy-preserving manner. Finally, numerical results are provided to demonstrate the effectiveness and efficiency of our scheme.©2021 IEEE	Parking management

Paper	Author	Date	Abstract	Application
FMFParking: Federated Matrix Factorization For Parking Lot Recommendation	C. Lu, Y. Fan, X. Wu and J. Zhang	2021	In recent years, artificial intelligence has developed rapidly and has been applied in various industries. In parking lot services, people have used data mining and analysis such as matrix factorization, decision tree to predict and recommend parking lots for users. However, with the rise of user privacy awareness and the promulgation of regulations, data acquisition and use rights have been restricted, which results in "data islands" and other issues. We construct a parking recommendation system framework of federated learning based on distributed encryption matrix factorization, called Federated Matrix Factorization for Parking Lot Recommendation" (FMFParking). In the system, we first build a framework based on the user distributed matrix factorization, and then design a homomorphic encryption federated learning with the framework of distributed matrix factorization. We have proved that this system can improve the security of private data while ensuring accuracy. In this system, each user's personal private data and their whereabouts to the parking lot are protected, which can encourage more users to join this system and provide more privacy-dimensional feature information. With the addition of parking lot data, we can increase the diversity and comprehensiveness of the data, which can make us build a more complete and personalized parking lot recommendation system. This article uses real Electronic Toll Collection (ETC) parking lot data and analyzes the experimental results.©2021 IEEE	Parking management
FedRC: A Federated Learning-Based Roadside Computing Paradigm Through the Facilitation of Internet of Drones	A. Islam and S. Y. Shin	2021	The modern era is filled with smart entities (e.g., smart vehicles) that have both sense and actuate capabilities. These entities can collect lots of data during their functional period and these data can be utilized for the wellbeing of citizens. However, these data are very sensitive raising issues like privacy. Moreover, network scarcity, bandwidth consumption, etc. can worsen the circumstance. Federated learning (FL), internet of drones (IoD), and dew computing (DC) are revolutionary technologies that can be engaged to mitigate the aforementioned challenges. An FL-based computing paradigm is initiated over the dew computing to process road-related data to bring efficiency in the applications (e.g., finding parking locations) utilizing IoD. An experimental environment is established containing a traffic dataset as a proof of concept. The experimental results exhibit the feasibility of the proposed scheme ©2021 IEEE	Road-related data collectioning by smart entities
FedCPF: An Efficient-Communication Federated Learning Approach for Vehicular Edge Computing in 6G Communication Networks	S. Liu, J. Yu, X. Deng and S. Wan	2021	The sixth-generation network (6G) is expected to achieve a fully connected world, which makes full use of a large amount of sensitive data. Federated Learning (FL) is an emerging distributed computing paradigm. In Vehicular Edge Computing (VEC), FL is used to protect consumer data privacy. However, using FL in VEC will lead to expensive communication overheads, thereby occupying regular communication resources. In the traditional FL, the massive communication rounds before convergence lead to enormous communication costs. Furthermore, in each communication round, many clients upload large quantity model parameters to the parameter server in the uplink communication phase, which increases communication overheads. Moreover, a few straggler links and clients may prolong training time in each round, which will decrease the efficiency of FL and potentially increase the communication costs. In this work, we propose an efficient-communication approach, which consists of three parts, including "Customized", "Partial", and "Flexible", known as FedCPF. FedCPF provides a customized local training strategy for vehicular clients to achieve convergence quickly through a constraint item within fewer communication rounds. Moreover, considering the uplink congestion, we introduce a partial client participation rule to avoid numerous vehicles uploading their updates simultaneously. Besides, regarding the diverse finishing time points of federated training, we present a flexible aggregation policy for valid updates by constraining the upload time. Experimental results show that FedCPF outperforms the traditional FedAVG algorithm in terms of testing accuracy and communication optimization in various FL settings. Compared with the baseline, FedCPF achieves efficient communication with faster convergence speed and improves test accuracy by 6.31% on average. In addition, the average communication optimization rate is improved by 2.15 times.©2021 IEEE	Vehicular efficient-communication

Paper	Author	Date	Abstract	Application
Charging Station Recommendation for Electric Vehicle Based on Federated Learning	Xiaohui Wang, Xiaokun Zheng and Xiao Liang	2021	At present, the usage of EV charging facilities is unbalanced. The accuracy of the charging station recommendation does not meet the demand. Due to the limitation of user privacy protection, charge point operators and vehicle enterprises cannot provide data to each other for joint analysis. Therefore, we proposed recommendation method of EV charge point based on federated learning. The federated factorization machine is implemented to make use of data features in both sides and cross features between them. We build the model by encrypted entity alignment, secure federated training and predicting. The experimental results show that the federated model improves the AUC of the model by 6% over those built with features only from the charge point operators. The model is superior to centralized LR-based and RF-based models. While the data does not need to leave the original platform, the model realizes the secure and precise federated charging point recommendation based on more comprehensive features.©2021 IOP Publishing Ltd	Charging station recommendation
A Blockchain based Federated Learning for Message Dissemination in Vehicular Networks	F. Ayaz, Z. Sheng, D. Tian and Y. L. Guan	2021	Message exchange among vehicles plays an important role in ensuring road safety. Emergency message dissemination is usually carried out by broadcasting. However, high vehicle density and mobility lead to challenges in message dissemination such as broadcasting storm and low probability of packet reception. This paper proposes a federated learning based blockchain-assisted message dissemination solution. Similar to the incentive-based Proof-of-Work consensus in blockchain, vehicles compete to become a relay node (miner) by processing the proposed Proofof-Federated-Learning (PoFL) consensus which is embedded in the smart contract of blockchain. Both theoretical and practical analysis of the proposed solution are provided. Specifically, the proposed blockchain based federated learning results in more vehicles uploading their models in a given time, which can potentially lead to a more accurate model in less time as compared to the same solution without using blockchain. It also outperforms other blockchain approaches in reducing 65.2% of time delay in consensus, improving at least 8.2% message delivery rate and preserving privacy of neighbour vehicle more efficiently. The economic model to incentivize vehicles participating in federated learning and message dissemination is further analysed using Stackelberg game. The analysis of asymptotic complexity proves PoFL as the most scalable solution compared to other consensus algorithms in vehicular networks.©2021 IEEE	Vehicular communication
Federated Learning for Object Detection in Autonomous Vehicles	D. Jallepalli, N. C. Ravikumar, P. V. Badarinath, S. Uchil and M. A. Suresh	2021	With the recent proliferation of Artificial Intelligence (AI), object detection is becoming increasingly ubiquitous. It is one of the key features of Autonomous Driving Systems. In current applications, object detection models are usually trained at a centralized location by collecting data from multiple sources. This raises concerns about data privacy among other issues. This paper addresses data privacy through a Federated Learning (FL) approach. FL architecture aims at preserving data privacy while maintaining performance by training the model in a decentralized manner. In this paper, we analyze how FL impacts the performance of object detection in a real-world traffic environment. We have constructed a prototype FL system and evaluated it on the KITTI Vision Benchmark 2D image dataset. In our prototype, object detection models are trained locally on a vehicle's dataset, and the resultant weights are securely aggregated, using symmetric encryption techniques during data transfer, at the global server to yield an improved model. The FL model converged at 68% mean average precision. We compared the performance of object detection using FL to the traditional deep learning approach and noticed significant difference between the two models.©2021 IEEE	Object detection

Paper	Author	Date	Abstract	Application
Privacy-Preserved Federated Learning for Autonomous Driving	Y. Li, X. Tao, X. Zhang, J. Liu and J. Xu	2021	In recent years, the privacy issue in Vehicular Edge Computing (VEC) has gained a lot of concern. The privacy problem is even more severe in autonomous driving business than the other businesses in VEC such as ordinary navigation. Federated learning (FL), which is a privacy-preserved strategy proposed by Google, has become a hot trend to solve the privacy problem in many fields including VEC. Therefore, we introduce FL into autonomous driving to preserve vehicular privacy by keeping original data in a local vehicle and sharing the training model parameter only with the help of MEC server. Moreover, different from the common assumption of honest MEC server and honest vehicle in former studies, we take the malicious MEC servers and malicious vehicles into account. First, we consider honest-but-curious MEC server and malicious vehicles and propose a traceable identity-based privacy preserving scheme to protect the vehicular message privacy where improved Dijk-Gentry-Halevi-Vaikutanathan (DGHV) algorithm is proposed and a blockchain-based Reputation-based Incentive Autonomous Driving Mechanism (RIADM) is adopted. Further, when the case comes to the non-credibility of both parties where semi-honest MEC server and malicious vehicles are considered, we propose an anonymous identity-based privacy preserving scheme to protect the identity privacy of vehicles with Zero-Knowledge Proof (ZKP). Based on the simulation of virtual autonomous driving based on real-world road images, it is verified that our proposes scheme can reduce 73.7 % training loss of autonomous driving, increase the accuracy to around 5.55 % while keeps effective privacy of message and identity under the threat of dishonest MEC server and vehicles.©2021 IEEE	Privacy problem in vehicular communication
Two-Layer Federated Learning With Heterogeneous Model Aggregation for 6G Supported Internet of Vehicles	X. Zhou, W. Liang, J. She, Z. Yan and K. I. -K. Wang	June 2021	The vision of the upcoming 6G technologies that have fast data rate, low latency, and ultra-dense network, draws great attentions to the Internet of Vehicles (IoV) and Vehicle-to-Everything (V2X) communication for intelligent transportation systems. There is an urgent need for distributed machine learning techniques that can take advantages of massive interconnected networks with explosive amount of heterogeneous data generated at the network edge. In this study, a two-layer federated learning model is proposed to take advantages of the distributed end-edge-cloud architecture typical in 6G environment, and to achieve a more efficient and more accurate learning while ensuring data privacy protection and reducing communication overheads. A novel multi-layer heterogeneous model selection and aggregation scheme is designed as a part of the federated learning process to better utilize the local and global contexts of individual vehicles and road side units (RSUs) in 6G supported vehicular networks. This context-aware distributed learning mechanism is then developed and applied to address intelligent object detection, which is one of the most critical challenges in modern intelligent transportation systems with autonomous vehicles. Evaluation results showed that the proposed method, which demonstrates a higher learning accuracy with better precision, recall and F1 score, outperforms other state-of-the-art methods under 6G network configuration by achieving faster convergence, and scales better with larger numbers of RSUs involved in the learning process.©2021 IEEE	Vehicular communication

Paper	Author	Date	Abstract	Application
Decentralized federated learning for extended sensing in 6G connected vehicles	Luca Barbieri, Stefano Savazzi, Mattia Brambilla, Monica Nicoli	2021	Research on smart connected vehicles has recently targeted the integration of vehicle-to-everything (V2X) networks with Machine Learning (ML) tools and distributed decision making. Among these convergent paradigms, Federated Learning (FL) allows the vehicles to train a deep ML model collaboratively, by exchanging model parameters (i.e., neural network weights and biases), rather than raw sensor data, via V2X links. Early FL approaches resorted to a server-client architecture, where a Parameter Server (PS) acts as edge device to orchestrate the learning process. Novel FL tools, on the other hand, target fog architectures where the model parameters are mutually shared by vehicles and synchronized in a distributed manner via consensus algorithms. These tools do not rely on the PS, but take advantage of low-latency V2X links. In line with this recent research direction, in this paper we investigate distributed FL methods for augmenting the capability of road user/object classification based on Lidar data. More specifically, we propose a new modular, decentralized approach to FL, referred to as consensus-driven FL (C-FL), suitable for PointNet compliant deep ML architectures and Lidar point cloud processing for road actor classification. The C-FL process is evaluated by simulating a realistic V2X network, based on the Collective Perception Service (CPS), for mutual sharing of the PointNet model parameters. The performance validation considers the impact of the degree of connectivity of the vehicular network, the benefits of continual learning over heterogeneous training data, convergence time and loss/accuracy tradeoffs. Experimental results indicate that C-FL complies with the extended sensors use cases for high levels of driving automation, it provides a low-latency training service, compared with existing distributed ML approaches, and it outperforms ego learning with minimal bandwidth usage. © 2021 Elsevier Inc.	6G connected vehicles
Model Aggregation Federated Learning Approach for Vehicular Traffic Forecasting	Savita Lonare, R. Bhramaram	2020	To make the Intelligent Transportation System (ITS) more efficient and robust, researchers are working hard. Ana traffic data helps ITS to be more helpful. Mobile phones are the prime source of traffic data. The vast availability and increased processing speed of mobile phones is making ITS more robust. Presently for traffic prediction, the e mobile user's data is accumulated at the central server. The information is then aggregated together to make prec In this approach, sensitive user data have the risk of privacy and security—massive user data uploading on the ser in latency.This paper proposes a decentralized approach for vehicular traffic prediction that allows 'selected' local organizations (clients) to train the model and share the trained model securely to the server. The selection of orga to participate in the training process is made by clustering algorithms. The server then aggregates the locally trai and shares the aggregated model to all the clients again	Traffic prediction
Road Vehicle Recognition in Monocular Images	M.A. Sotelo, J. Nuevo, L.M. Bergasa, M. Ocana, I. Parra, D. Fernandez	2005	This paper describes a monocular vision-based Vehicle Recognition System in which the basic components of road vehicles are first located in the image and then combined with a SVM-based classifier. The challenge is to use a single camera as input. This poses the problem of vehicle detection and recognition in real, cluttered road images. A distributed learning approach is proposed in order to better deal with vehicle variability, illumination conditions, partial occlusions and rotations. The vehicle searching area in the image is constrained to the limits of the lanes, which are determined by the road lane markings. By doing so, the rate of false positive detections is largely decreased. A large database containing thousands of vehicle examples extracted from real road images has been created for learning purposes. We present and discuss the results achieved up to date. ©2005 IEEE	Vehicle recognition system

Paper	Author	Date	Abstract	Application
A Learning Automata Based Algorithm For Solving Capacitated Vehicle Routing Problem	Mir Mohammad Alipour	2012	This paper presents an approximate algorithm based on distributed learning automata for solving capacitated vehicle routing problem. The vehicle routing problem (VRP) is an NP-hard problem and capacitated vehicle routing problem variant (CVRP) is considered here. This problem is one of the NP-hard problems and for this reason many approximate algorithms have been designed for solving it. Distributed learning automata that is a general searching tool and is a solving tool for variety of NP-complete problems, is used to solve this problem and tested on fourteen benchmark problems. Our results were compared to the best known results. The results of comparison have shown the efficiency of the proposed algorithm. (c) 2012 International Journal of Computer Science Issue	Vehicle routing problem
Anomaly Removal for Vehicle Energy Consumption in Federated Learning	G. Lin, X. Zhu, J. Wang and J. Xiao	2021	Federated learning is a distributed deep learning method that enables parallel and distributed learning of data on multiple participants, without the need to centrally store it. In intelligent transportation system, it is impractical to gather the vehicle data from the edge devices due to data privacy concerns or network bandwidth limitation. Hence, combining with federated learning to train vehicle data processing models has become one of the popular solutions. However, such computing paradigm is subject to threats posed by malicious and abnormal nodes that greatly reduces the computing power of the neural network when performing calculations in a distributed manner. In this paper, we use the Vehicle Energy Dataset to simulate distributed vehicle data. Based on these data, we propose an unsupervised anomaly removal and neural network model based on federated learning to solve the problem of outlier data on vehicle equipment and analyze the effect of speed on fuel consumption. The results show that with the proposed anomaly removal strategy, MAE and MSE of the trained network are 29% and 36% lower than those without anomaly removal, respectively. ©2021 IEEE	Vehicle energy consumption
Content-based Vehicle Selection and Resource Allocation for Federated Learning in IoV	S. Wang, F. Liu and H. Xia	2021	In order to use datasets collected from multiple vehicles to train a machine learning model while ensuring vehicle user privacy, federal learning framework was introduced into the Internet of Vehicles. Federated learning is a distributed learning framework. Under the federated learning framework, the packet error rate and wireless bandwidth have a great influence on the global model training process because the vehicle exchanges model parameters with the central server through the wireless channel. With limited bandwidth, the central server needs to select a more appropriate subset of vehicles candidates to participate in federated learning. In this paper, image classification is taken as a typical application in the Internet of vehicles. The dataset contents of different vehicles are different, and the selection of different subsets of vehicles will affect the accuracy and convergence rate of the global model. Therefore, an algorithm of vehicle selection and wireless resource allocation based on dataset content is proposed. Vehicles selection and wireless resource allocation are designed as an optimization problem by joint considerations of vehicle computing resource, datasets, and wireless resources with the goal of maximizing loss function decay of the global model. Finally, simulation with the CIFAR-10 dataset verifies that the vehicle selection and resource allocation algorithm based on the dataset content is superior to the baseline methods in terms of model accuracy and convergence rate.©2021 IEEE	vehicle selection and wireless resource allocation

Paper	Author	Date	Abstract	Application
A Review on Vehicle to Vehicle Communication Protocols in VANETs	Adil Mudasir Malla,Ravi Kant Sahu	2013	Vehicular Ad hoc Networks is a form or type of mobile ad-hoc network to provide communication among nearby vehicles and nearby fixed equipments or roadside units for improving efficiency and safety of transportation. Even though it possesses characteristics of high node mobility and fast topology changes but still it can provide wide variety of services, ranging from safety related warning message system for improved navigation mechanism as well as information and entertainment applications. In this paper, we have studied various mechanisms or techniques along their comparison and limitation which were used to handle the communication challenges like congestion, delay, collision, redundancy while propagating emergency warning messages in Vehicular Ad hoc Networks (VANeTs), as it is the case where if these communication challenges are not controlled may result in traffic accidents leading to human loss. © 2013, IJARCSSE All Rights Reserved	Vehicle to vehicle Communication
Reinforcement Learning Scheduler for Vehicle-to-Vehicle Communications Outside Coverage	T. Şahin, R. Khalili, M. Boban and A. Wolisz	2018	Radio resources in vehicle-to-vehicle (V2V) communication can be scheduled either by a centralized scheduler residing in the network (e.g., a base station in case of cellular systems) or a distributed scheduler, where the resources are autonomously selected by the vehicles. The former approach yields a considerably higher resource utilization in case the network coverage is uninterrupted. However, in case of intermittent or-of-coverage, due to not having input from centralized scheduler, vehicles need to revert to distributed scheduling.Motivated by recent advances in reinforcement learning (RL), we investigate whether a centralized learning scheduler can be taught to efficiently pre-assign the resources to vehicles for-of-coverage V2V communication. Specifically, we use the actor-critic RL algorithm to train the centralized scheduler to provide non-interfering resources to vehicles before they enter the-of-coverage area.Our initial results show that a RL-based scheduler can achieve performance as good as or better than the state-of-art distributed scheduler, often outperforming it. Furthermore, the learning process completes within a reasonable time (ranging from a few hundred to a few thousand epochs), thus making the RL-based scheduler a promising solution for V2V communications with intermittent network coverage. ©2018 IEEE	Vehicle to vehicle Communication
A Distributed Anomaly Detection System for In-Vehicle Network Using HTM	C. Wang, Z. Zhao, L. Gong, L. Zhu, Z. Liu and X. Cheng	2018	With the development of 5G and Internet of Vehicles technology, the possibility of remote wireless attack on an in-vehicle network has been proven by security researchers. Anomaly detection technology can effectively alleviate the security threat, as the first line of security defense. Based on this, this paper proposes a distributed anomaly detection system using hierarchical temporal memory (HTM) to enhance the security of a vehicular controller area network bus. The HTM model can predict the flow data in real time, which depends on the state of the previous learning. In addition, we improved the abnormal score mechanism to evaluate the prediction. We manually synthesized field modification and replay attack in data field. Compared with recurrent neural networks and hidden Markov model detection models, the results show that the distributed anomaly detection system based on HTM networks achieves better performance in the area under receiver operating characteristic curve score, precision, and recall. ©2028 IEEE	Distributed anomaly detection system

Paper	Author	Date	Abstract	Application
A Dispersed Federated Learning Framework for 6G-Enabled Autonomous Driving Cars	L. U. Khan, Y. K. Tun, M. Alsenwi, M. Imran, Z. Han, C. S. Hong	2021	Sixth-Generation (6G)-based Internet of Everything applications (e.g. autonomous driving cars) have witnessed a remarkable interest. Autonomous driving cars using federated learning (FL) has the ability to enable different smart services. Although FL implements distributed machine learning model training without the requirement to move the data of devices to a centralized server, it its own implementation challenges such as robustness, centralized server security, communication resources constraints, and privacy leakage due to the capability of a malicious aggregation server to infer sensitive information of end-devices. To address the aforementioned limitations, a dispersed federated learning (DFL) framework for autonomous driving cars is proposed to offer robust, communication resource-efficient, and privacy-aware learning. A mixed-integer non-linear (MINLP) optimization problem is formulated to jointly minimize the loss in federated learning model accuracy due to packet errors and transmission latency. Due to the NP-hard and non-convex nature of the formulated MINLP problem, we propose the Block Successive Upper-bound Minimization (BSUM) based solution. Furthermore, the performance comparison of the proposed scheme with three baseline schemes has been carried out. Extensive numerical results are provided to show the validity of the proposed BSUM-based scheme. ©2021 IEEE	Vehicle networking
Distributed Learning Agents in Urban Traffic Control	Eduardo Camponogara, Werner Kraus Jr.	2003	Automatic learning techniques stand as promising tools to respond to the need of higher efficiency of traffic network, even more so at times of mounting pressure from economic and energy markets. To this end, this paper looks into the operation of a traffic network with distributed, intelligent agents. In particular, it casts the task of operating a traffic network as a distributed, stochastic game in which the agents solve reinforcement-learning problems. Results from computational experiments show that these agents can yield substantial gains with respect to the performance achieved by two other control policies for traffic lights. The paper ends with an outline of future research to deploy machine-learning technology in real-world traffic networks. © Springer-Verlag Berlin Heidelberg 2003	Traffic prediction

2 WP2 - Identifying Attack Scenarios and Countermeasures

The objective of this work package is the identification of attack scenarios aimed in particular at disclosing and manipulating user data or model logic, as well as merging user data with other databases. To do so, we first identify relevant transmission paths and the transmitted data from the use cases obtained in chapter 1. We then identify attack scenarios on transmission channels and involved devices. Lastly, we determine different trust scenarios between the involved entities.

2.1 Attacks Against Federated Learning

In this section, we have a closer look at attacks against federated learning architectures.

2.1.1 Gradient inversion attack

This art of an attack is about data recovery after eavesdropping on the communication between the global server and participants. State-of-the-art gradient inversion attacks are stronger because they can make two assumptions about batch normalization statistics or/and private labels. [Huang et al., 2021b]

2.1.2 Poisoning attacks

Poisoning attack based on generative adversarial nets (GAN)

An attacker sends identical data samples as benign participants by replicating their samples using GAN. Then, the attacker manipulates his data in further learning rounds. [Zhang et al., 2019]

Data Poisoning attacks

Data Poisoning can be classified as a clean-label attack and a dirty-label (label-flipping attack). By the label flipping attack or targeted data poisoning attack a central server of federated learning receives manipulated training data from malicious participants. These participants transfer incorrect data to disturb aggregated information and afterward the classification mechanism itself. [Tolpegin et al., 2020]

Distributed poisoning attacks

In a distributed poisoning attack several attackers coorperate with each other. The attackers falsify data in local models to disturb samples to the learning process. [Cao et al., 2019]

Model Poisoning attacks

It is an attack on local models by modifying their parameters. The aims are similar to the data poisoning attack. [Mammen, 2021]

2.1.3 Backdoor

Nuding and Mayer [2020] show in their work about traffic sign classification that federated learning is vulnerable against backdoor attacks. In their work they found that for sequential training, the later in the training process, the fewer were non-adversarial nodes able to decrease the effectiveness of the backdoor.

2.1.4 Standard falsified information attack

Malicious participants provide Road Side Units (RSU) with falsified real-time updates whilst entering a study zone of RSU multiple times in a short period. [Al Mallah et al., 2021]

2.1.5 Sybil attacks

Malicious participant transfers their incorrect samples under different IDs created to duplicate false data. [Al Mallah et al., 2021]

2.1.6 Model replacement attacks

Model replacement attacks are called single-shot backdoor attacks, because of being implemented only in one round. On the contrary, a multiple-shot attack occurs in more than one round. [Zhao et al., 2021]

2.1.7 Membership inference attacks

An attacker aims to gather information about the learning process and its dynamics. To implement it, the attacker requests a global model to predict the presence of certain samples. [Pustozerova and Mayer, 2020]

2.2 Defense methods

In this section, we have a closer look at possible defense methods, identified in literature.

2.2.1 Defense methods against gradient attacks

Noisy gradients

Gaussian and Laplacian noise distributions with a certain variance range help to disturb accuracy of the image recovery. [Zhu et al., 2019]

Gradient Compression and Sparsification

Zhu et al. [2019] show in their work that pruning ratio of gradients which is around twenty percent violates accuracy of the recovered images. Because DLG (Deep leakage from gradients) is hard to implement, if gradients are compressed.

MixUp

MixUp is a defense method that encrypts images to provide secure communication between the global server and participants. [Huang et al., 2021b]

InstaHide

The creation of InstaHide was inspired by Mixup. This new defense scheme has two versions to carry out the encryption of inputs: Inter-InstaHide and Intra-InstaHide. Inter-InstaHide uses images from the public dataset to mix up with an image, whereas Intra-InstaHide from the private dataset. [Huang et al., 2021b]

2.2.2 Defense methods against poisoning attacks

Federated learning system aggregator

FL system aggregator identifies malicious data by using clustering, which excludes suspicious and dissimilar training samples. [Tolpegin et al., 2020]

Defense against Model Poisoning Attacks

Mammen [2021] propose a defense method that uses the Error rate and the Loss function to identify manipulated local models.

Reliable worker selection scheme

A defense method that uses reputation as a reliability metric to identify malicious participants.[Kang et al., 2020]

Foolsgold

Fung et al. [2018] introduce the novel defense scheme against targeted poisoning attacks, such as sybil-based label-flipping and backdoor poisoning attacks.It is a new approach to identify Sybils by observing the contribution similarity of federated learning's participants.

Sniper

Cao et al. [2019] demonstrate in their work a defense against distributed poisoning attack. Sniper is a new scheme for identifying poisoned local models.

Spectral anomaly detection framework

The novel spectral anomaly detection framework is useful for unsupervised as well as semi-supervised FL. The global server defines a threshold using information received from each client. Then, eliminates samples of malicious participants by applying the threshold. This framework helps against targeted model poisoning attacks and also against Byzantine attacks. [Li et al., 2020]

Quantum-Based Federated Learning Framework

An Optimized Quantum-Based federated learning framework for defending against adversarial attacks in intelligent transportation systems. [Yamany et al., 2021]

CONTRA

Awan et al. [2021] propose a novel defense scheme CONTRA against poisoning attacks. This generic scheme analyzes clients' updates by examining their alignment level each training round. Clients with an increasing alignment level will be detected as suspicious.

2.2.3 Defense against backdoor attacks

Local and Central Differential Privacy (LDP/CDP)

Naseri et al. [2020] propose in their work Local and Central Differential Privacy (LDP/CDP) as a defense mechanism against backdoor attacks. Both of them can defend against backdoor attacks and white-box membership inference attacks, but they are not robust to property inference attacks.

FederatedReverse

FederatedReverse is a defense technique against image backdoor attacks, which has 4 parts. Firstly, a reverse trigger will be generated for each label by participants with help of reverse engineering. Secondly, the central server creates global reverse triggers by using global reverse trigger generation. Thirdly, in order to find out malicious samples, the central server performs outlier detection. In the fourth part, the model repair liquidates all samples of an attacker. [Zhao et al., 2021]

Norm thresholding of updates

The norm thresholding of updates has a global model setting to not accept samples, which are above a certain threshold. [Sun et al., 2019]

Neuron Pruning Methods

Wu et al. [2020] compare two neuron pruning methods. The first method uses a ranking vote, which helps to maintain security and privacy because there is no need of knowing about real values. Clients average their value and compose ranking vote in the last convolutional layer. The second method provides more protection for the clients' privacy. Clients assign neurons as ”0” (to be pruned) or ”1” (not to be pruned) by using a pruning rate provided by the global server. After that, the global server provides a majority vote for all neurons. In the end, by ranking vote as well as by majority vote the server removes all neurons till the point, where the accuracy on the validation dataset is under a certain threshold.

Table 2.1 – Risk Evaluation

Gradient	Membership Inference	Data Poisoning	Gradient Attack
Traffic Flow	-	+	-
Energy Demand	(+)	+	+
Eco-routing	+	+	+
Vehicular Object Detection	(+)	+	(+)
Parking Space Estimation	+	(+)	+

Table 2.2 – Identifying Countermeasures

Gradient	Data Poisoning	Membership Inference
Local Differential privacy (LDP) Gradient perturbation with additive noise Gradient squeezing with controlled local training iterations	Reliable worker selection e.g. Blockchain Poisoning detection e.g. error rate, loss function	Differential privacy (LDP/CDP) Homomorphic encryption Secure multiparty computation

3 WP3 - Assessing Countermeasures

The objective of work package 3 is the identification and subsequent assessment of suitable countermeasures against the in WP2 defined attack scenarios for the use cases developed in WP1. In WP3.1, the identified countermeasures are discussed and compared, under consideration of the trust scenarios identified in WP2. In WP3.2 we construct a designated test network to evaluate the effectiveness and efficiency of the identified countermeasures. We determine the potentially achievable level of data protection for the identified use cases. The feasibility of countermeasures in the individual scenarios and the added value through distributed learning is discussed.

3.1 WP3.1 - Identification of countermeasures

In this section we present the different trust scenarios and several possible attacks. We then also take a look at the possible de-identification techniques.

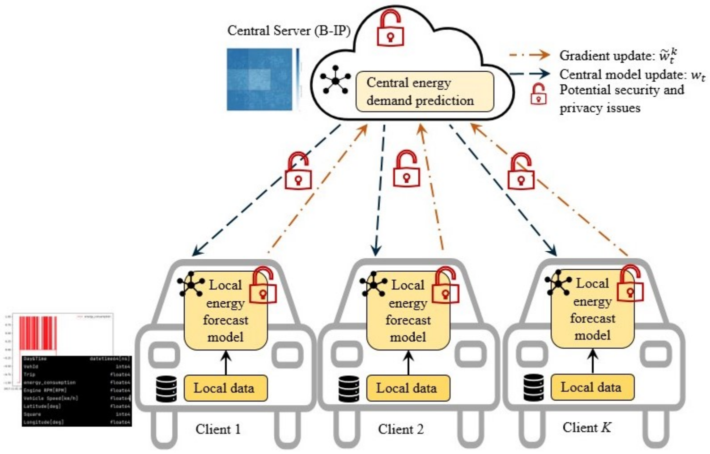
3.1.1 Federated learning trust scenarios and countermeasures

The general architecture of the FL LSTM model is shown in Figure 3.1. There we also indicate the possible places an attack can happen with an open lock.

Data poisoning attack

In this scenario one or more vehicles are not trustworthy. These attacking vehicles are poisoning the central model with gradient updates from mislabeled data. In general two different model poisoning attacks exist. First, the attacker aims to decrease the accuracy of the overall model (untargeted). In our example, the

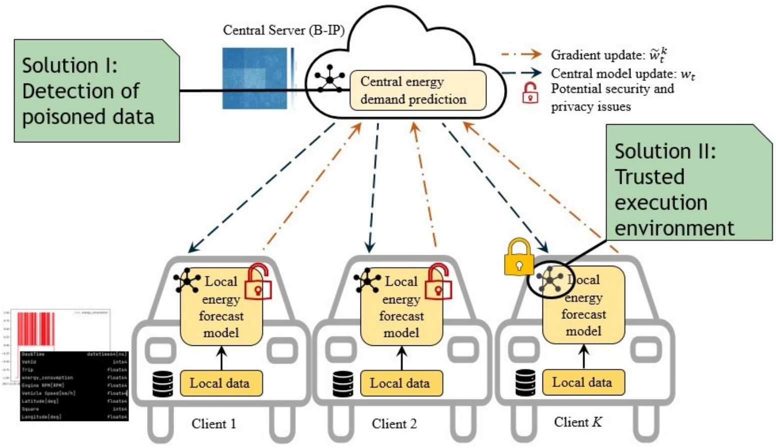
Figure 3.1 – Architecture and threats



Central Server (B-IP)
Central energy demand prediction
Gradient update: $\tilde{w}_t^k$
Central model update: w_t
Potential security and privacy issues
Local energy forecast model
Local data
Client 1
Client 2
Client K

accuracy for all squares would be reduced. Second, the attacker aims to achieve a misclassification in a certain class (targeted). An example is the wrong prediction of the energy consumption at a specific time of the day in a certain square. Based on the results in WP2 we summarize backdoor attacks (poisoning several rounds), model replacement attacks (single shot backdoor), sybil attacks (several IDs to poison) and label flipping as model poisoning attacks.

Figure 3.2 – Countermeasures. Data Poisoning

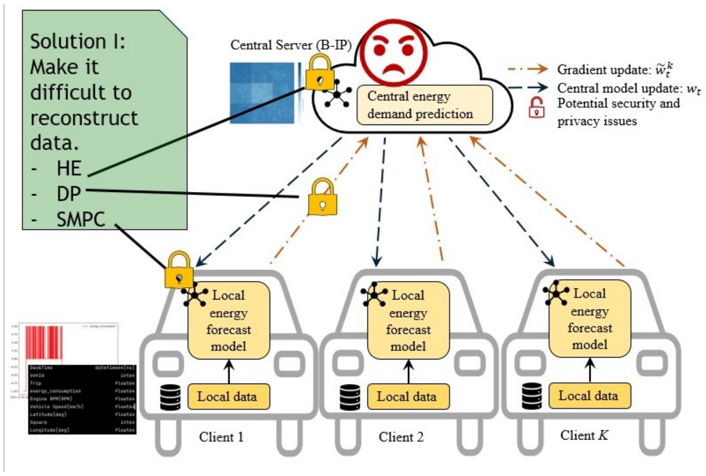


Central Server (B-IP)
Solution I: Detection of poisoned data
Central energy demand prediction
Gradient update: $\tilde{w}_t^k$
Central model update: w_t
Potential security and privacy issues
Solution II: Trusted execution environment
Local energy forecast model
Local data
Client 1
Client 2
Client K

We have identified the following solutions for our scenario to mitigate model poisoning. First, poisoned data can be detection by e.g., comparing the loss functions. Second, a reliability management can be implemented that only allows trusted participants, e.g., using blockchain technology. The major drawback of this method is that rare events may be not learned because they are rejected from the detection algorithm. Third, trusted execution environments (TEE) can be used to ensure that the code and data are not changed. Nevertheless, unintended data can still be created e.g., by running vehicles on a test stand.

Privacy implications: Although this attack is a serious threat for federated learning the direct privacy impact of this attack for a certain user is low. But in case a trust management is included the question of fairness and ethical correct detection of untrustworthy participants has to be taken carefully into account.

Figure 3.3 – Countermeasures. Membership Inference Attack - Server



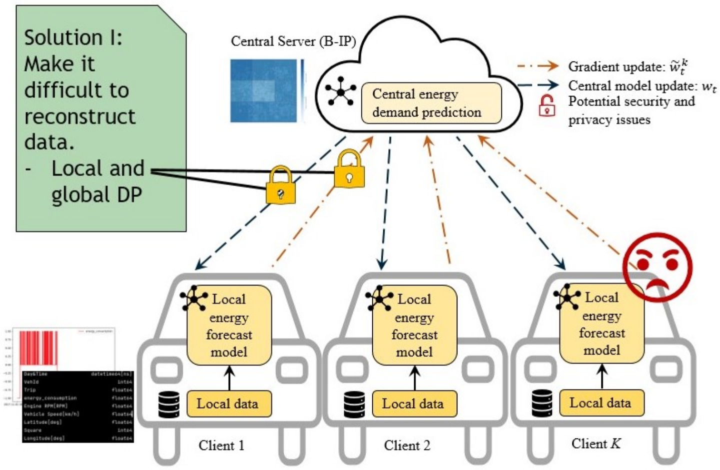
```
DayGTime                     datetime64[ns]
VehId                                 int64
Trip                                float64
energy_consumption                  float64
Engine RPM[RPM]                     float64
Vehicle Speed[km/h]                 float64
Latitude[deg]                       float64
Square                                int64
Longitude[deg]                      float64
```

Membership inference attacks

In a membership inference attack the attacker aims to observe the output of a certain model to make conclusions about the training data, e.g., by reconstructing data used to train a local model. Attacker can be the central server attacking the model of a certain vehicle, another vehicle attacking the global model or a third person (see Figure 3.4).

The solutions to mitigate this attack all base on the principle to make the reconstruction of a model more difficult. First, this can be done by using HE. We evaluate HE as very strong but also very complex to be implemented. With SMPC the gradients can be protected by using SMPC to perform the aggregation

Figure 3.4 – Countermeasures. Membership Inference Attack - Server


Solution I:
Make it difficult to reconstruct data.
- Local and global DP
Central Server (B-IP)
Central energy demand prediction
Gradient update: $\widetilde{w}_t^k$
Central model update: w_t
Potential security and privacy issues
Local energy forecast model
Local data
Client 1
Local energy forecast model
Local data
Client 2
Local energy forecast model
Local data
Client K
energy_consumption

```
Day&Time                 datetime64[ns]
VehId                             int64
Trip                            float64
energy_consumption              float64
Engine RPM[RPM]                 float64
Vehicle Speed[km/h]             float64
Latitude[deg]                   float64
Square                            int64
Longitude[deg]                  float64
```

algorithm between the vehicles before sending the gradients to the central server. This only works for an aggregation between vehicles but not for the gradients of the central server. Central and local differential Privacy (LDP/CDP) can be used to protect the vehicles gradients as well as the model's gradients. This technique is strong but has an impact on accuracy.

Privacy implications: In our scenario, an attacker could receive insights in the past GPS data and therefore reveal the motion profile of a certain vehicle. Also the energy consumption of a vehicle can be interesting for an attacker because it contains information about the driving profile.

Gradient attacks

In a gradient the attacker aims to recover data after eavesdropping on the communication between the global server and the participants. State-of-the-art gradient inversion attacks are strong because they can make assumptions about batch normalization statistics and also about the private labels [Huang et al., 2021b] (see Figure 3.5).

The mitigation strategy against gradient attacks also uses de-identification techniques to increase the difficulty of eavesdropping on the communication between the global server and the participants. This again includes HE, SMPC, LDP and CDP with the above described advantages and disadvantages.

Figure 3.5 – Countermeasures. Gradient Attack

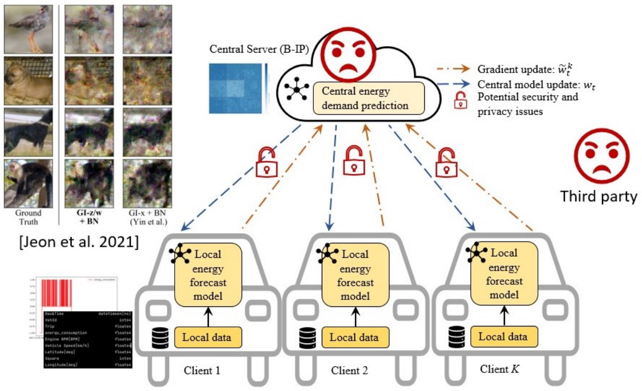
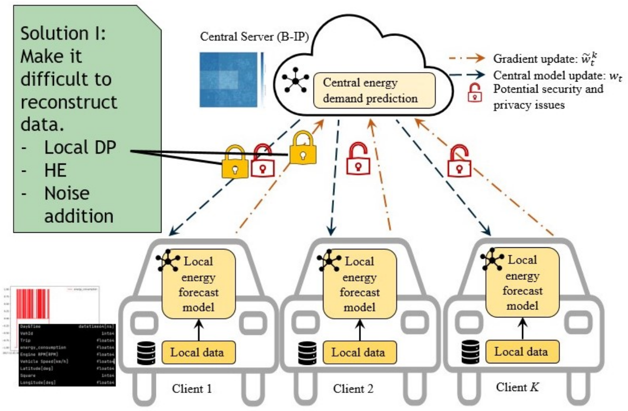

Ground Truth
GI-z/w + BN
GI-x + BN (Yin et al.)
[Jeon et al. 2021]
Central Server (B-IP)
Central energy demand prediction
Gradient update: $\tilde{w}_t^k$
Central model update: w_t
Potential security and privacy issues
Third party
Local energy forecast model
Local data
Client 1
Local energy forecast model
Local data
Client 2
Local energy forecast model
Local data
Client K

Figure 3.6 – Countermeasures. Gradient Attack


Solution I:
Make it difficult to reconstruct data.
- Local DP
- HE
- Noise addition
Central Server (B-IP)
Central energy demand prediction
Gradient update: $\tilde{w}_t^k$
Central model update: w_t
Potential security and privacy issues
Local energy forecast model
Local data
Client 1
Local energy forecast model
Local data
Client 2
Local energy forecast model
Local data
Client K

Privacy implications: In our scenario, the attacker can start a model inversion attack after reveling the gradients. Again, this will then reveal the past GPS data and therefore reveal the motion profile of a certain vehicle. Also the energy consumption of a vehicle can be interesting for an attacker because it contains information about the driving profile.

Summary table for FL extensions

In Table 3.1 we summarize the countermeasures that can be used in combination with federated learning, e.g. differential privacy, secure multiparty computation, homomorphic encryption.

Table 3.1 – Identifying Countermeasures

Gradient	Data Poisoning	Membership inference
Local Differential privacy (LDP) Gradient perturbation with additive noise Gradient squeezing with controlled local training iterations	Reliable worker selection e.g. Blockchain Poisoning detection e.g. error rate, loss function Trusted execution environment (TEE)	Differential privacy (LDP/CDP) Homomorphic encryption Secure multiparty-computation

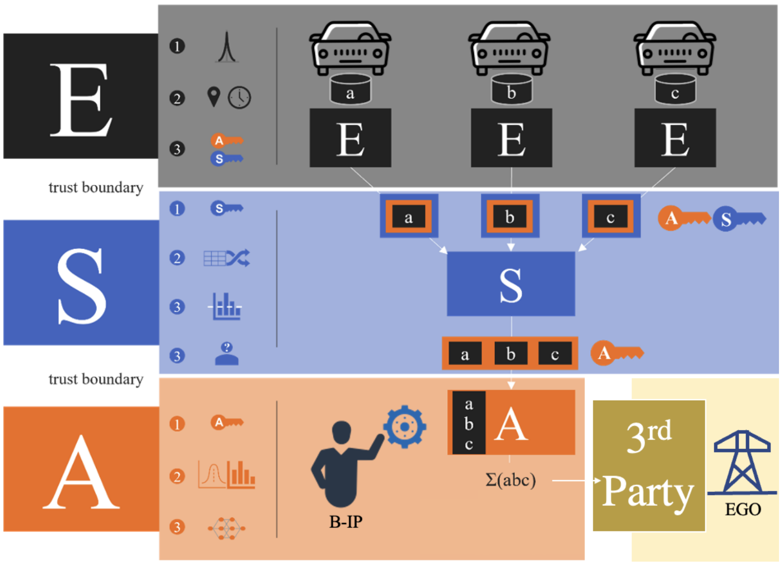
3.1.2 ESA trust scenarios and countermeasures

In this section we describe problems that can arise within the ESA architecture.

Trust boundaries

In the ESA architecture several trust boundaries exist. These trust boundaries can be corrupted and thereby data of a single vehicle can be linked and revealed. The trust boundaries are depicted in Figure 3.7.

Figure 3.7 – ESA Architecture adapted to the energy demand prediction use case (compare Bittau et al.) (2017)



A mitigation strategy is e.g. proposed by Erlingsson et al. [2020] who introduce a shuffling layer of K independent shuffler what gives the driver the opportunity to choose a shuffler that matches their privacy needs.

Membership inference attacks in ESA

Again, in a membership inference attack the attacker tries to reconstruct the output of a model that was built on the data to make conclusions about the training data, e.g., by reconstruct data used to train a local model. In the ESA architecture, this kind of attack can for instance be performed by an analyser.

This attack can be mitigated by the architecture itself if the correct values for the batchsize are chosen:

Setting the correct ϵ: The ϵ value is also known as privacy budget and should be aimed to be kept as low as possible e.g. 0.05. A very low ϵ value is a good indicator for a good anonymization but should not be treated as a guarantee. Also a visualization of the anonymized data that is then compared to the real data or a test by using a membership inference attack should be considered.

Setting the correct threshold The threshold is relevant to built the group a single vehicle can hide in. If it is set to low, a vehicle cannot be hidden in the crwod what has a negative effect on the privacy. In general, [Bittau et al., 2017] recommend a batch size of 20. A to high batchsize is in general not bad for the privacy but can have a negative impact on the performance and accuracy of the model.

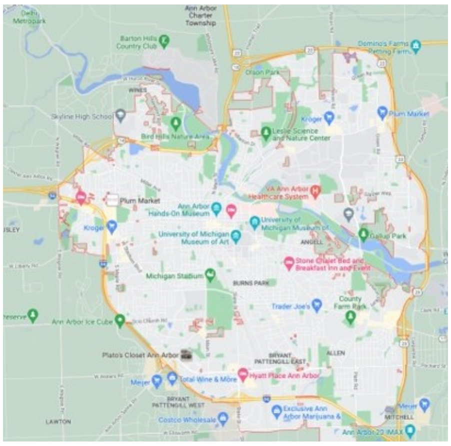
3.2 WP3.2 - Demonstrators

In this section we present the FL approach and the ESA demonstrators. During the implementation phase of the FL demonstrator, the shortage of data of vehicles that are driving at the same time and the small trip duration impeded the implementation so strongly that we decided to utilize the ESA architecture instead of the FL demonstrator for this use-case. This solution provides us to mitigate the obstacle of the lack of data and provides better results. Therefore, we will provide the approach of the FL architecture and then continue explaining the ESA architecture. Finally, we compare both technologies based on the existing results and give an outlook on future implementations.

3.2.1 Data preparation

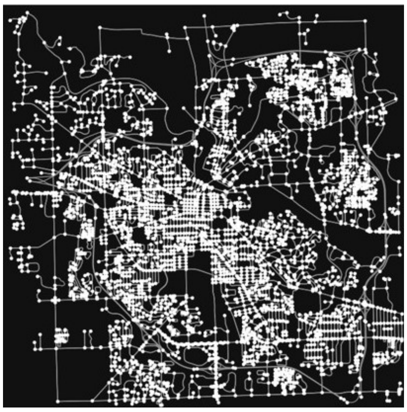
Oh et al. [2019] created an extensive dataset, titled VED (Vehicle Energy Dataset), which includes information about the fuel and energy consumption of 383 personal vehicles in Ann Arbor, Michigan, USA (see Figure 3.8 and Figure 3.10).

Figure 3.8 – Google Maps



Source: Google Maps

Figure 3.9 – Graph based on GPS locations



Source: Generated from OpenStreetMap data with OSMX by Sonja Rieger GUF

Onboard OBD-II loggers from Nov, 2017 to Nov, 2018 were used to investigate GPS trajectories, fuel, energy, speed, and auxiliary power usage data of vehicles. In total, the dataset captures 374,000 miles, driven by vehicles in varying conditions, ranging from highways to downtown areas (see Table 3.2).

Table 3.2 – Vehicle Energy Dataset (VED)

Personal vehicles (Total)	Gasoline vehicles	HEVs	PHEV/EVs
383	264	92	27

The dataset is divided into two parts: Dynamic Data and Static Data. Each part captures 383 vehicles from a different perspective.

The Dynamic Data includes a week's worth of time-stamped naturalistic driving records. The table of the Dynamic Data represents data in the following columns: DayNum, VehId, Trip, Timestamp(ms), Latitude[deg], Longitude[deg], Vehicle Speed[km/h], MAF[g/sec], Engine RPM[RPM], Absolute Load[Percent], Outside Air Temperature[DegC], Fuel Rate[L/hr], Air Conditioning Power[kW], Air Conditioning Power[Watts], Heater Power[Watts], HV Battery Current[A], HV Battery SOC[Percent], HV Battery Voltage[V], Short Term Fuel Trim Bank 1[Percent], Short Term Fuel Trim Bank 2[Percent], Long Term Fuel Trim Bank 1[Percent], Long Term Fuel Trim Bank 2[Percent].

The Static Data on the other hand captures vehicle parameters of all 383 vehicles. The dataset, inter alia, includes data from three 2013 Nissan Leaf, which are pure EV vehicles. The following columns are represented in the table of the Static Data: VehId, EngineType, Vehicle Class, Engine Configuration and Displacement Transmission, Drive Wheels, Generalized Weight[lb] (see Table 3.3 and Table 3.4).

Table 3.3 – Data preparation. Selected Data

Day Num	Veh Id	Trip	Timestamp (ms)	Latitude [deg]	Longitude [deg]	Vehicle Speed [km/h]	MAF [g/sec]	Engine RPM [RPM]	Absolute Load [%]	OAT [DegC]	Fuel Rate [L/hr]	Air Conditioning Power [kW]	Air Conditioning Power [Watts]
148.81195262	7	1328	0	42.3159052778	-83.734226667	69	6.2399997112	1143	14.9019613266	10	NaN	NaN	NaN
148.81195262	7	1328	200	42.3159052778	-83.734226667	69	6.2600002888	1118	15.2941179276	10	NaN	NaN	NaN

Table 3.4 – Data preparation. Selected Data

Heater Power [Watts]	HV Battery Current[A]	HV Battery SOC[%]	HV Battery Voltage[V]	Short Term Fuel Trim Bank 1[%]	Short Term Fuel Trim Bank 2[%]	Long Term Fuel Trim Bank 1[%]	Long Term Fuel Trim Bank 2[%]
NaN	NaN	NaN	NaN	-4.6875	0.78125	-2.34375	-2.34375
NaN	NaN	NaN	NaN	-4.6875	0.78125	-2.34375	-2.34375

Table 3.5 – EV Data Used

VehId	Trip	DayNum	Latitude	Longitude	Vehicle Speed	Outside Air Temperature	Battery Voltage	Battery Current	Energy Consumption
1	371	82.45...	42.24...	-83.76...	33.6...	30.25...	-441...	319.5...	9.48

Table 3.6 – ICE Data Used

VehId	Trip	DayNum	Latitude	Longitude	Vehicle Speed	Engine RPM	Energy Consumption
1	371	82.45...	42.24...	-83.76...	33.6...	845	9.48

Preparation of GPS data

To predict the energy demand for the city in this model, we separate the city into 9 squares (see figures 3.11). All squares are of the same size and each vehicle is mapped to exactly one square at a time t. The sum of the energy consumption at a time t in a square is defined as the actual energy consumption in a specific square.

Figure 3.10 – Graph based on GPS locations

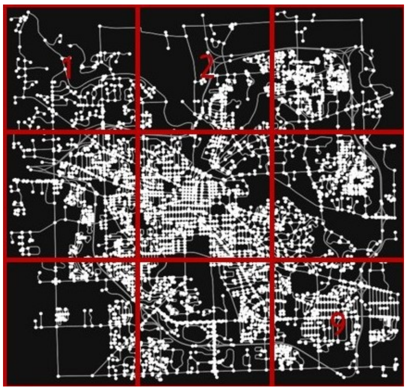
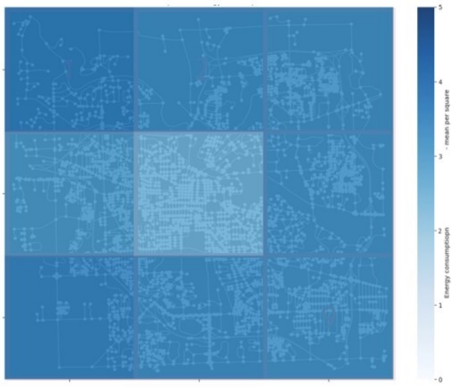


Figure 3.11 – Graph based on GPS locations



Preparation of energy consumption

The multiplication of battery voltage (“HV Battery Current[A]”) and current (“HV Battery Voltage[V]”) has as a result the target feature of energy consumption for EV. The outcome can be both positive and negative, because of battery current. It has positive values when charging during driving using the technique of regenerative breaking. But when the energy is consumed the battery current has negative values. The battery voltage has only positive values. Oh et al. [2019]

$$EnergyConsumption = HVBatteryCurrent[A] * HVBatteryVoltage[V] \quad (3.1)$$

With A = Ampere and V = Volt

An approach data based on a dummy code can be applied to calculate the remaining fuel rates using the given OBD-II data. [1] The fuel rate for ICE vehicles can be calculated by the multiplication of a correction rate and mass air flow, if the MAF is given. The correction rate considers the air composition of fuel and air. For its calculation there are following information is needed: Short-Term Fuel Trim Bank 1 (STFT) and Long-Term Fuel Trim Bank 1 (LTFT) and the air fuel ratio (AFR), which standard value is 14.7 and can be used for gasoline vehicles. Rimpas et al. [2020]

$$FuelRate[g/s] = massairflow(MAF) * correction \quad (3.2)$$

$$correction = (1 + STFT * \frac{1}{100} + \frac{LTFT}{100}) * \frac{1}{100} AFR \quad (3.3)$$

STFT=short term fuel trim; LTFT=long term fuel trim; AFR=air fuel ratio

The average of STFT trims 1 and 2 is used for the calculation, if they both are given. When the MAF is in the measuring unit of g/s, then there is a need of its conversion to the measuring unit of the fuel rate given [2] in other data records. Meseguer et al. [2015] have proposed the following formula:

$$FuelRate[l/h] = \frac{(MAF * 3600)}{airfuelratio} * fueldensitygasoline \quad (3.4)$$

[1] OBD-II scanners are used to perform an emission test. They show better results in the minimizing the “emission” produced by vehicles. Tim Miller: ”OBD2: The definite guide about on-board diagnostics II”

[2] Meseguer et al. [2015] use the term of “Fuel Flow” instead of “Fuel Rate”. To keep consistency the original term used in Oh et al. [2019] is maintained.

with an AFR of 14.7 as standard and a fuel density of gasoline with 820.

To calculate mass air flow, if there is not any MAF data available, the following information is needed: the absolute load ("Absolute Load [%]"), the air resistance ("pair"), the displacement of the engine (Displacement(eng)), the engine speed in rotation per minute ("Engine RPM [RPM]"), the correction rate and a standard air resistance of 1.184 [3]. Meseguer et al. [2015]

$$massairflow(MAF) = 1.184[g/l]*AbsLoad*\frac{1}{100}*\Big(Displacement_{eng})[l]*\frac{1}{2}*RPM[RPM]*\frac{1}{60} \tag{3.5}$$

Oh et al. [2019] are using the formula also in this form:

$$massairflow(MAF) = \frac{AbsLoad}{100}*1.184[g/l]*\Big(Displacement_{eng})[l]*\frac{EngineRPM[RPM]}{120} \tag{3.6}$$

From the static data field "Engine Configuration" can be derived the displacement of the engine, because it was not directly given. In liters is given the information about the displacement is before the "L". It is given the exemplary value if "4-FI 1.5L". There are also other data records, which can not be considered for the prediction model for the reason that they did not offer any of the fields used for fuel consumption above and therefore do not offer a target feature.

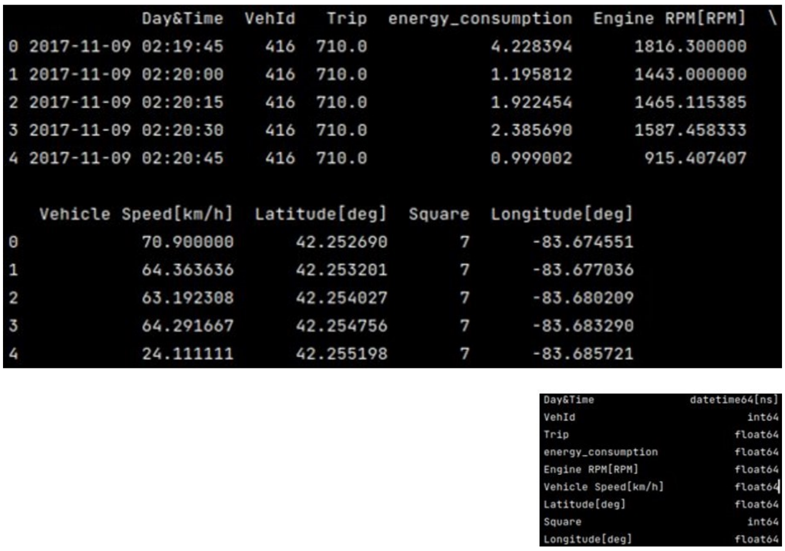
Data preparation for Federated Learning

Final input data for LSTM: In Figure 3.12 we provide an overview of the first 5 rows of the data of Vehicle 416, Trip 710. In this test run, we have aggregated the time of the trip to 15 seconds. The variable square is the assignment of the GPS location to a square in the map.

Test run local LSTM: Before building a federated network we want to better understand the data prediction on a single vehicle. Therefore, we built a model for only a single trip (Vehicle 416, Trip 710). This model loss of the test run

[3] For the calculation of the MAF another source than the dataset proposing paper has been used. Therefore the following formula is using a standard air resistance of 1.184.

Figure 3.12 – Selected Data



```
             Day&Time  VehId   Trip  energy_consumption  Engine RPM[RPM]  \
0 2017-11-09 02:19:45    416  710.0            4.228394      1816.300000
1 2017-11-09 02:20:00    416  710.0            1.195812      1443.000000
2 2017-11-09 02:20:15    416  710.0            1.922454      1465.115385
3 2017-11-09 02:20:30    416  710.0            2.385690      1587.458333
4 2017-11-09 02:20:45    416  710.0            0.999002       915.407407

   Vehicle Speed[km/h]  Latitude[deg]  Square  Longitude[deg]
0            70.900000      42.252690       7      -83.674551
1            64.363636      42.253201       7      -83.677036
2            63.192308      42.254027       7      -83.680209
3            64.291667      42.254756       7      -83.683290
4            24.111111      42.255198       7      -83.685721
```

```
Day&Time                datetime64[ns]
VehId                            int64
Trip                           float64
energy_consumption             float64
Engine RPM[RPM]                float64
Vehicle Speed[km/h]            float64
Latitude[deg]                  float64
Square                           int64
Longitude[deg]                 float64
```

is shown in Figure 3.13. Figure 3.14 provides an overview of the actual energy consumption during this trip.

Figure 3.13 – Model Loss

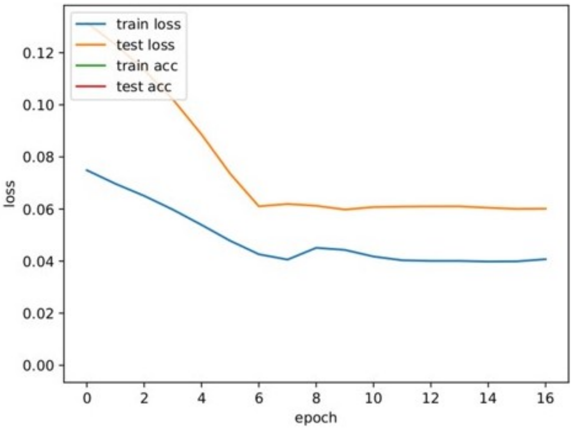
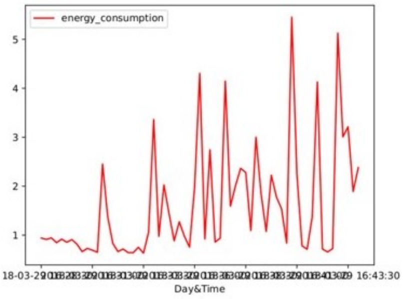
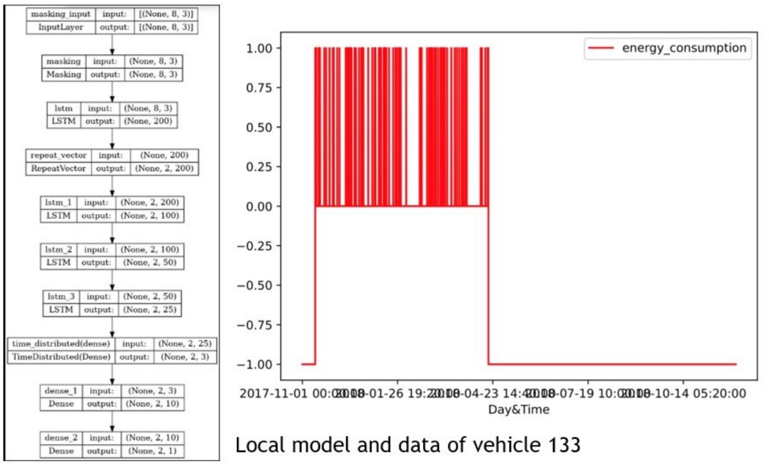


Figure 3.14 – Energy Consumption



In Figure 3.15 we show the architecture of the lstm in the left and all trips of Vehicle 416 on the right. In this architecture, we have implemented a masking layer to not learn time series data that is marked with a −1. The input data for this model is a complete time series from beginning of the experiment to the end. Missing time stamps are marked with −1. On the right of Figure 3.15 we can see that the vehcile was not driving in the second half of the experiment.

Figure 3.15 – Missings in time Series



Data preparation for ESA

For the tests with the ESA architecture, we have split the dataset into EV and ICE vehicle and show the used data in Table 3.7 and Table 3.8. Crowd IDs were assigned based on the time and the geo location to be later used in the shuffling step. We have seperated the map into 100 squares and assigned square IDs accordingly.

Table 3.7 – EV data used in the prototype

VehId	Trip	DayNum	Latitude	Longitude	Vehicle Speed	Outside Air Temperature	Battery Voltage	Battery Current	Energy Consumption
1	371	82.45...	42.24...	-83.76...	33.6...	30.25...	-441...	319.5...	9.48

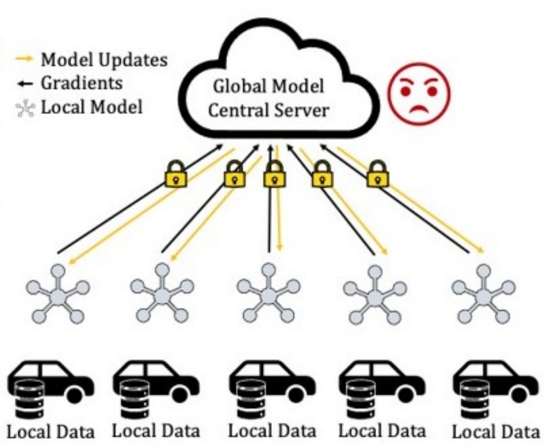
Table 3.8 – ICE data used in the prototype

VehId	Trip	DayNum	Latitude	Longitude	Vehicle Speed	Engine RPM	Energy Consumption
1	371	82.45...	42.24...	-83.76...	33.6	845	9.48

3.2.2 Federated Learning Pre-Test of Demonstrator

In this section we aimed to construct a test network for different tasks to evaluate the attack scenarios identified in section 3.1.2. The architecture that we planned can be found in Figure 3.16. The pre-tests did not show promising results. The main reason for this is a lack of data that is required to train the LSTM and a lack of existing prototypes that provide parameter tuning for an LSTM in a FL environment.

Figure 3.16 – De-identifying transmitted data



From the conducted pre-tests we are able to draw some conclusions for the federated model. First, to build a complete time series for learning, masking layers have to be implemented. It is important to make sure that not only masked data is used for training. Second, the aggregation to 15 seconds might be too strong since there is a high fluctuation in energy consumption as shown in Figure 3.14. The aggregation might have a negative impact on model performance. Third, with regard to privacy, GPS data can have a positive impact on accuracy and should therefore not be transformed to a square ID before the learning.

3.2.3 Encode, Shuffle, Analyse (ESA)

In the following we will introduce the Encode, Shuffle, Analyse (ESA) architecture that uses Local Differential Privacy. There are three threats that arise as a result of the ESA architecture:

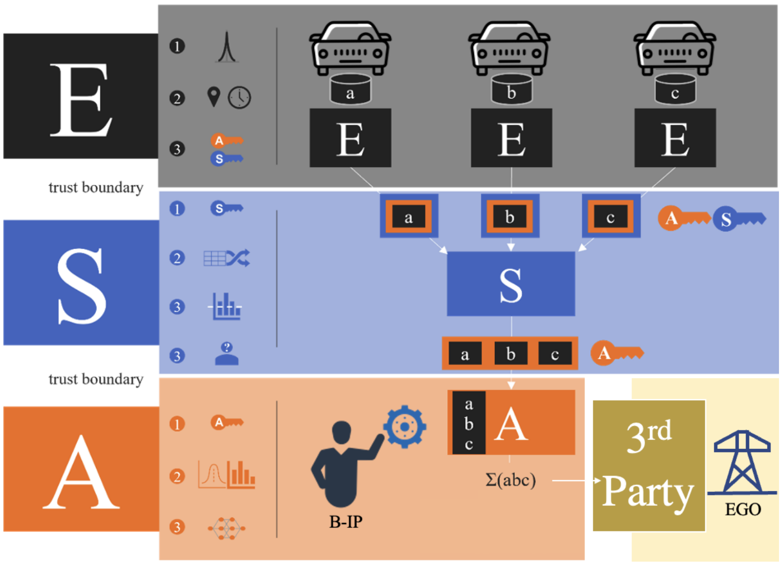
Table 3.9 – Threats

Location determination:	Breaking perturbation with additional data helps B-IP to determine GPS location.
Vehicle tracking and track localization:	GPS information can be observed by a malicious B-IP.
Linkability and Profiling:	Reidentification of vehicles and creation of profiles is possible with the help of personal data.

Approach

The ESA architecture is presented in Figure 3.17 in the way it was introduced by Bittau et al. [2017]. Figure 3.18 represents all required steps in more detail. Further we will discuss the steps separately.

Figure 3.17 – ESA Architecture (cf. Bittau et al.)(2017)



As shown in Figure 3.19, the architecture consists of Encoder, Shuffler, and Analyzer. There are two layers of encryption and corresponding analyzer and shuffler keys.

There are five steps for the Encoder part: data preparation (e.g.missing values, outlier), local differential privacy with Laplacian Noise, crowd ID for shuffling,

Figure 3.18 – ESA Steps in detail

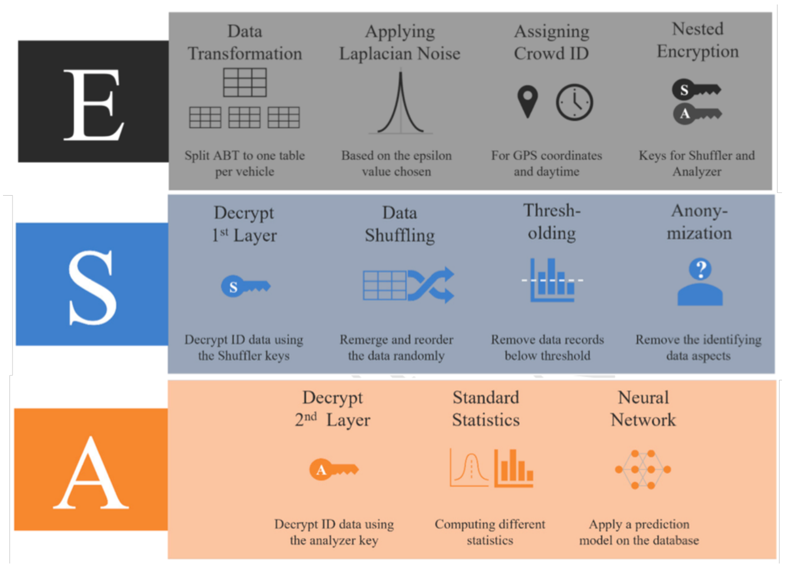
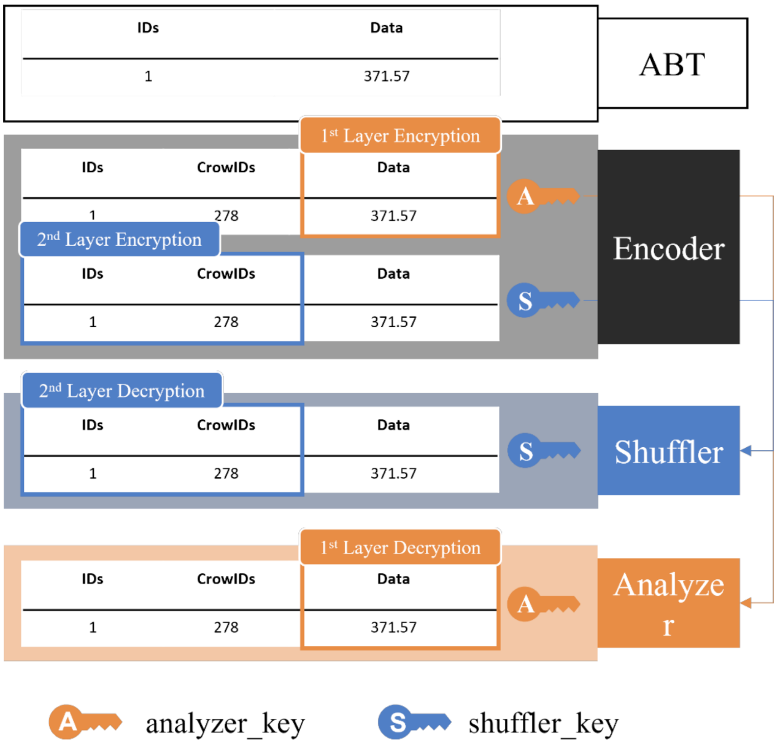

E
Data Transformation
Split ABT to one table per vehicle
Applying Laplacian Noise
Based on the epsilon value chosen
Assigning Crowd ID
For GPS coordinates and daytime
Nested Encryption
S
A
Keys for Shuffler and Analyzer
S
Decrypt 1st Layer
S
Decrypt ID data using the Shuffler keys
Data Shuffling
Remerge and reorder the data randomly
Thresh-olding
Remove data records below threshold
Anony-mization
Remove the identifying data aspects
A
Decrypt 2nd Layer
A
Decrypt ID data using the analyzer key
Standard Statistics
Computing different statistics
Neural Network
Apply a prediction model on the database

Figure 3.19 – ESA nested encryption



nested encryption (Layer A and Layer S), and sending data to the Shuffler.

The Shuffler contains four following steps: decryption of the layer S, shuffling by randomly reordering data records per crowd (size 20), removing metadata (VehId", "Trip", and "DayNum"), sending the data to the Analyser.

For the Analyser there are only two steps to implement: decryption of the layer A and analysing data, such as energy consumption and Neural Network.

Results

In order to understand real impact of the ESA Architecture and the quality of the data we will observe the data as well as after applying this architecture, and, in addition, without implementing ESA. For the latter case we constructed a baseline model that is indicated with an ϵ of 0 for EV and ICE. In the following we discuss results only for EVs.

First, we analyze square ID energy consumption for the EVs on the heat map for $\epsilon = 0.05$ and $\epsilon = 0.5$ with the threshold of 20 for the shuffling. We observe in Figure 3.20, that much viewer squares are provided after applying the threshold and ϵ value.

As a second step, we analyze the computed energy consumption per hour, which is shown in Figure 3.21. The barplot with an ϵ of 5 has a partially analogous structure as our baseline model, meanwhile the barplot with an ϵ of 0.5 appears to be strongly dissimilar with the model. At the same same time all three barplots have identical maximum value.

Then we analyze the MSE of our neural network. In Figure 3.22 it can be seen that in comparison with the baseline model the MSE for an ϵ of 0.5 converges at much later periods. For an ϵ of 5 it occurs earlier but the training curve is much higher than the validation.

Furthermore, we consider data of the ICE vehicle. We show the mean of consumption per square on the heat map and the energy consumption over daytime. Repeatedly we compare the baseline model of ICE vehicles with the threshold for $\epsilon = 0.5$ to a value of 20 for the shuffling. From Figures 3.23 and 3.24 it is visible that the dataset is much larger for ICE. Additionally, heat maps and barplots appear to be very similar. We assume that the quantity of data has big impact

Figure 3.20 – Square ID energy consumption EV for ϵ values 0, 0.5, 5

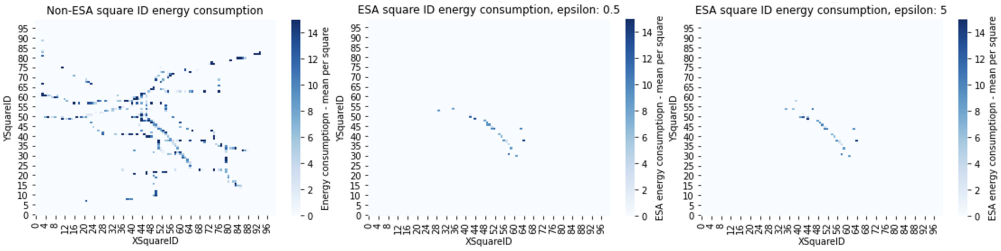


Figure 3.21 – Consumption over daytime EV for ϵ values 0, 0.05, 0.5

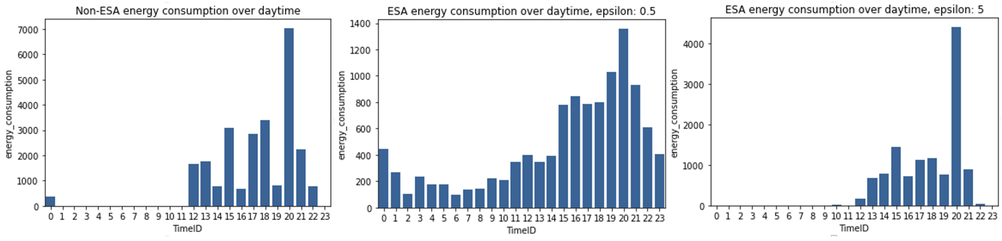


Figure 3.22 – MSE EV of ϵ values 0, 0.05, 0.5)

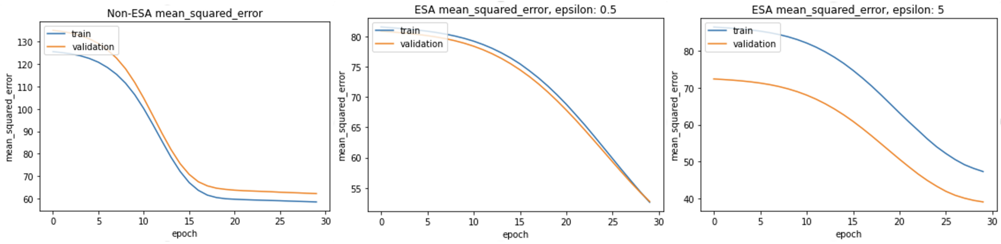


Figure 3.23 – Square ID energy consumption ICE for ϵ values 0, 0.5

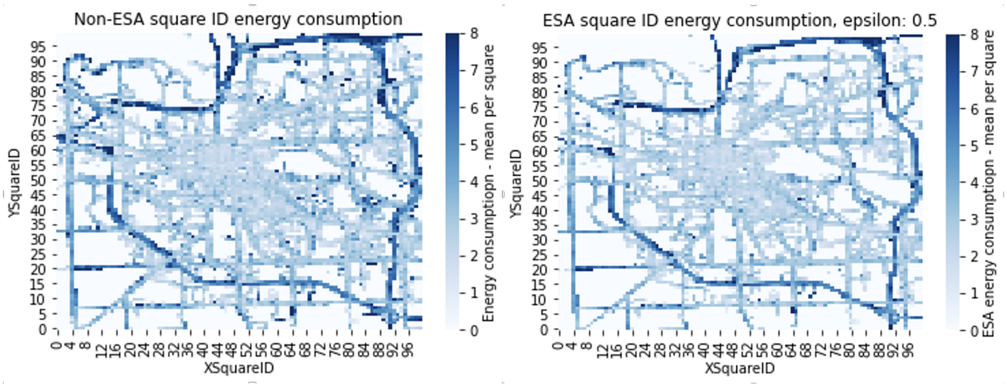
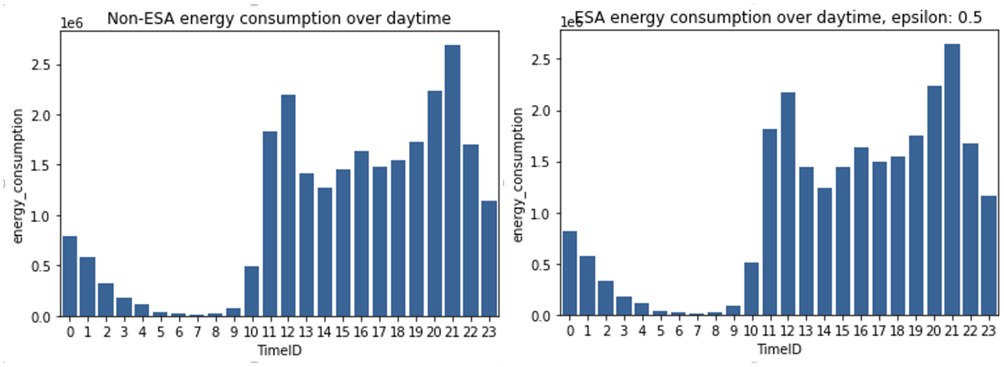


Figure 3.24 – Consumption over daytime ICE for ϵ values 0, 0.5



on the quality. Therefore, the analysis of the EV data could have shown better results, if there would be more information available.

3.3 Comparison ESA and FL

In this section we compare the application of both technologies in the scenario of energy demand prediction. In Table 3.10 we compare ESA and FL based on different factors and issues, e.g. involved entities, achieved level of privacy, efficiency evaluation, and implementation hurdles identified by the prototype.

Entities

Both architectures exhibit a trusted party, therefore it is worthwile to compare their trust implications. While a malicious central server will be able to learn data

Table 3.10 – Comparison ESA and FL

	ESA	FL
Entities	• Analyser (B-IP): Does not need to be honest since data is shuffled. • Shuffler: Additional entity that removes data	• Central Server (B-IP): Often required to be honest. If not LDP required to de-identify gradients what reduces the accuracy
Privacy	Strong (low Epsilon) if shuffler trustworthy and enough data is available	Tbd (assumed to be good for trustworthy B-IP but a privacy accuracy trade-off exists)
Efficiency	Only efficient if enough data available (min 20 per crowd)	For an efficient analysis enough data is required.
Hurdles	• Data availability / number of participating vehicles • Trust into shuffler • Key exchange	• Data availability • Synchronization of vehicles, especially between different vehicle/chip generations • GPU power in vehicle
Type of ML	Supervised and unsupervised learning	Mostly used for supervised learning

and reconstruct models about all participants this is not possible for a malicious shuffler in the ESA architecture since the inner encryption can only be decrypted by the analyser. Nevertheless, a collaboration of analyser and shuffler is possible and has to be prevented.

Privacy

For the federated learning architecture a strong level of privacy can only be achieved if the central server is trusted or the transmission paths are protected with e.g. differential privacy. For the ESA architecture we could show that for $\epsilon = 0.05$ and $\epsilon = 0.5$ and a batch size of 20 a good level of privacy can be achieved.

Efficiency

For both methods, it has been shown that they only work, if enough data is available. For the FL architecture we did not find efficient results with the provided data. For the ESA architecture we were able to show that with $\epsilon = 0.05$ and $\epsilon = 0.5$ and a batch size of 20 for the shuffling usable results with a good level of privacy can be achieved.

Hurdles

As already mentioned, the implementation hurdles for FL are trust in the central server, data availability and the synchronization of vehicles. Especially the synchronization has become an issue since first, the different vehicle types cannot be compared and second, the trip duration is more important than expected at the beginning.

Also for the ESA architecture some hurdles for implementation have been identified. Here, also an entity, the shuffler exist that has to be trusted to not collaborate with the analyser.

Type of ML

During the building of the prototypes we have found rarely examples for unsupervised learning with federated learning. We asses the technology mostly be used for classification problems. We expect more complex implementations in future that can also better deal with unsupervised learning problems.

Since the ESA architecture relies on a central analyser, all different types of machine learning can be used here, e.g. natural language processing.

3.4 Review of previous results

In this section we will repeat and enrich the aggregated results of the de-identification techniques with the findings of this report. In Table 3.11 we repeat the overview of the attribute-evaluation for all de-identification techniques based on the previous report [Löbner et al., 2021, Rannenberg et al., 2021]. These results were derived from the de-identification technique specific analysis.

Table 3.11 – Aggregated results of de-identification techniques [Löbner et al., 2021, Rannenberg et al., 2021]

	HE	MPC	Distr. DP	FL	K-anon.
Protective effect	⊕ High	⊕ High	⊕ High	⊕ High[2]	⊙ Medium
Complexity	⊖ High	⊖ High	⊖ High	⊖ High	⊕ Low
Runtime	⊖ High	⊖ High	⊖ High	⊖ High	⊙ Medium
Degree of maturity	⊙ Medium	⊙ Medium	⊕ High	⊙ Medium	⊕ High
Implement. effort	⊖ High	⊖ High	⊖ High	⊖ High	⊕ Low
Monetary cost	⊙ Medium	⊖ High	⊖ High	⊖ High	⊕ Low
Time blur	⊖ High	⊙ Medium	⊕ Low	⊙ Medium	⊙ Medium
Location obfus.	⊕ Low	⊖ High	⊕ Low	⊕ Low	⊖ High
Processing speed	⊖ Low	⊖ Low	⊕ High	⊕ High	⊙ Medium
Time delay	⊙ Medium	⊖ High	⊙ Medium	⊙ Medium	⊙ Medium
Aggregated data	Yes	Yes	Yes	Yes	Yes
Truthfulness	Yes	No	Yes [1]	Yes	No

[1] (No for GPS)

[2] (Only if central server is trustworthy)

Protective effect (overall level of privacy that can be achieved through the de-identification technique): For the distributed differential privacy that is represented by the ESA Architecture, we still evaluate the protective effect to be high since we could show that a good protection can be achieved in our scenario . For FL the evaluation of the protective effect has to be investigated in more details. In case of a scenario where the central server is trustworthy we still evaluate the protective effect to be high. In case that the central server is malicious we lower our evaluation to medium which we indicate with a footnote in the mapping table. A good protective effect can still be achieved if the technology is extended by another de-identification technique.

Complexity (overall complexity to develop, implement and maintain a particular solution): For both the ESA Architecture and FL the complexity of the technology is still high. But especially for FL it has to be ensured that the models can run on different platforms or vehicle generations which adds complexity. Since the ESA Architecture relies on a centralised analysis, data or platform issues can also be handled centrally.

Runtime (time that the overall solution for a use case needs to perform all necessary tasks): In our use case we have not investigated the execution time due to the scientific set up of entities.

Degree of maturity (scientific and commercial advancement of a de-identification technique): To implement the ESA Architecture was no big issue since all the steps that have assigned tasks, e.g. shuffler and analyser are well defined in the literature. Also extensions exist already. For FL a lot of academic literature exists already, but the implementation can be complex since an alignment of hardware and software is necessary among different devices, if the architecture should not be recreated from scratch. Therefore, we stick to the previous evaluation.

Implementation effort (overall effort that needs to be taken to implement the solution for a specific use case): Especially with creating extra entities such as the shuffler in the ESA Architecture and the central server in FL extra implementation effort has to be expected. Also not all hardware in the FL architecture might be able to train a network which can generate issues among different generations of devices. Therefore, we stick to the previous evaluation.

Monetary cost (cost of development and procurement of all necessary hard- and software): We stick to the previous evaluation since extra entities come with extra costs.

Time blur (degree to which data loses information that are related to a specific time point): Based on our results we would not change the evaluation but want to add that data in the ESA Architecture can get lost if the batch size is not reached or can be changed by the local differential privacy layer.

Location obfuscation (degree of obfuscation applied in a specific scenario, e.g. aggregated on a street, city or kilometer basis): The location obfuscation can be set based on the use case. A good model should show a low location obfuscation for both ESA and FL.

Processing speed (execution time of the de-identification technique itself): The execution time fits the use case.

Time delay (delay with which data is reported and can be acted upon): Both technologies need to build batches. In the ESA Architecture they are used to hide vehicles. In FL the batches are used to improve the model. Therefore, we stick to the evaluation of medium.

Aggregated data (describes a state in which data that is gathered during a use case is aggregated and thus a loss of information in the data occurs): In both cases data is aggregated. Therefore, the evaluation is correct.

Truthfulness (describes whether input data and output data are equal when using a de-identification technique): In our scenario, truthfulness of data does not exist in the ESA Architecture since we used differential privacy on all data and the metadata is removed. This can change according to the scenario. With FL, trustfulness does exist if a local model is used but it does not exist for the gradients of the trained model, especially if the gradients are protected by e.g. differential privacy.

4 Outlook

In this work we have enriched the findings from the current status of academic literature on two de-identification techniques: Federated Learning (FL) and the Encode, Shuffle, Analyse (ESA) architecture that implements distributed differential privacy in a combination with an additional shuffler step to break the linkability of participating entities.

This report focuses on the construction of a demonstrator to identify implementation hurdles and to identify possible countermeasures. This is done using vehicle location data for energy demand prediction.

Our findings demonstrate that although FL and the ESA architecture were, in the beginning and based on theoretical knowledge, both suitable for the use-case, they exhibit different characteristics. While we initially planned to implement an FL approach, we found during our pre-test that this technology does not work considering the requirements of our energy demand prediction use-case. It is important to mention that this does not mean that FL should not be used in vehicular use cases at all, but it holds for this use case. This provides valuable insights in the current implementation issues and strengths and weaknesses of the current state in practice of FL.

With the implementation of the ESA architecture we enriched the results of the previous evaluation of de-identification techniques by the findings of this demonstrator and were able to validate almost all of them. We conclude that both models require substantial amounts of data of many vehicles at the same time. Although we were able to show that a strong level of de-identification of data is possible with the ESA architecture we also find that the accuracy increases with the amount of data provided.

In future we aim to validate also the other proposed de-identification techniques in the framework from [Rannenberg et al., 2021] to obtain a validated framework for all de-identification techniques.

Bibliography

Study on the technical evaluation of de-identification procedures for personal data in the automotive sector. *Universitätsbibliothek Johann Christian Senckenberg*, May 2021. doi: http://dx.doi.org/10.21248/gups.63413. URL `http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/63413`.

Ranwa Al Mallah, Godwin Badu-Marfo, and Bilal Farooq. Cybersecurity threats in connected and automated vehicles based federated learning systems. In *2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*, pages 13–18. IEEE, 2021.

Sana Awan, Bo Luo, and Fengjun Li. Contra: Defending against poisoning attacks in federated learning. In *European Symposium on Research in Computer Security*, pages 455–475. Springer, 2021.

Andrea Bittau, Úlfar Erlingsson, Petros Maniatis, Ilya Mironov, Ananth Raghunathan, David Lie, Mitch Rudominer, Ushasree Kode, Julien Tinnes, and Bernhard Seefeld. PROCHLO: Strong Privacy for Analytics in the Crowd. In *SOSP 2017 - Proceedings of the 26th ACM Symposium on Operating Systems Principles*, 2017. ISBN 9781450350853. doi: 10.1145/3132747.3132769.

Di Cao, Shan Chang, Zhijian Lin, Guohua Liu, and Donghong Sun. Understanding distributed poisoning attack in federated learning. In *2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*, pages 233–239, 2019. doi: 10.1109/ICPADS47876.2019.00042.

Ahmet M. Elbir and Sinem Coleri. Federated Learning for Vehicular Networks, 2020. ISSN 23318422.

Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Shuang Song, Kunal Talwar, and Abhradeep Thakurta. Encode, shuffle, analyze privacy revisited: Formalizations and empirical evaluation. *arXiv preprint arXiv:2001.03618*, 2020.

Clement Fung, Chris JM Yoon, and Ivan Beschastnikh. Mitigating sybils in federated learning poisoning. *arXiv preprint arXiv:1808.04866*, 2018.

Xumin Huang, Peichun Li, Rong Yu, Yuan Wu, Kan Xie, and Shengli Xie. Fedparking: a federated learning based parking space estimation with parked vehicle assisted edge computing. *IEEE Transactions on Vehicular Technology*, 70(9): 9355–9368, 2021a.

Yangsibo Huang, Samyak Gupta, Zhao Song, Kai Li, and Sanjeev Arora. Evaluating gradient inversion attacks and defenses in federated learning. *Advances in Neural Information Processing Systems*, 34, 2021b.

Jiawen Kang, Zehui Xiong, Dusit Niyato, Yuze Zou, Yang Zhang, and Mohsen Guizani. Reliable federated learning for mobile networks. *IEEE Wireless Communications*, 27(2):72–80, 2020.

Xiangjie Kong, Kailai Wang, Mingliang Hou, Xinyu Hao, Guojiang Shen, Xin Chen, and Feng Xia. A federated learning-based license plate recognition scheme for 5g-enabled internet of vehicles. *IEEE Transactions on Industrial Informatics*, 2021.

Suyi Li, Yong Cheng, Wei Wang, Yang Liu, and Tianjian Chen. Learning to detect malicious clients for robust federated learning. *arXiv preprint arXiv:2002.00211*, 2020.

Yi Liu, JQ James, Jiawen Kang, Dusit Niyato, and Shuyu Zhang. Privacy-preserving traffic flow prediction: A federated learning approach. *IEEE Internet of Things Journal*, 7(8):7751–7763, 2020.

Sascha Löbner, Frédéric Tronnier, Sebastian Pape, and Kai Rannenberg. Comparison of de-identification techniques for privacy preserving data analysis in vehicular data sharing. In *Computer Science in Cars Symposium*, pages 1–11, 2021.

Chiaoyu Lu, Yushun Fan, Xing Wu, and Junqi Zhang. Fmfparking: Federated matrix factorization for parking lot recommendation. In *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, pages 131–136. IEEE, 2021.

Priyanka Mary Mammen. Federated learning: Opportunities and challenges. *arXiv preprint arXiv:2101.05428*, 2021.

Javier E Meseguer, Carlos T Calafate, Juan Carlos Cano, and Pietro Manzoni. Assessing the impact of driving behavior on instantaneous fuel consumption. In *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)*, pages 443–448. IEEE, 2015.

Mohammad Naseri, Jamie Hayes, and Emiliano De Cristofaro. Toward robustness and privacy in federated learning: Experimenting with local and central differential privacy. *arXiv e-prints*, pages arXiv–2009, 2020.

Florian Nuding and Rudolf Mayer. Poisoning attacks in federated learning: An evaluation on traffic sign classification. In *Proceedings of the Tenth ACM Conference on Data and Application Security and Privacy*, pages 168–170, 2020.

GS Oh, David J Leblanc, and Huei Peng. Vehicle energy dataset (ved), a large-scale dataset for vehicle energy consumption research. *arXiv preprint arXiv:1905.02081*, 2019.

Anastasia Pustozerova and Rudolf Mayer. Information leaks in federated learning. In *Proceedings of the Network and Distributed System Security Symposium*, 2020.

Kai Rannenberg, Sebastian Pape, Frederic Tronnier, and Sascha Löbner. Study on the technical evaluation of de-identification procedures for personal data in the automotive sector. Technical report, Technical Report. Goethe University Frankfurt. https://doi. org/10.21248 . . . , 2021.

Dimitrios Rimpas, Andreas Papadakis, and Maria Samarakou. Obd-ii sensor diagnostics for monitoring vehicle operation and consumption. *Energy Reports*, 6: 55–63, 2020.

Yuris Mulya Saputra, Dinh Thai Hoang, Diep N Nguyen, Eryk Dutkiewicz, Markus Dominik Mueck, and Srikathyayani Srikanteswara. Energy demand prediction with federated learning for electric vehicle networks. In *2019 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2019.

Ziteng Sun, Peter Kairouz, Ananda Theertha Suresh, and H Brendan McMahan. Can you really backdoor federated learning? *arXiv preprint arXiv:1911.07963*, 2019.

Vale Tolpegin, Stacey Truex, Mehmet Emre Gursoy, and Ling Liu. Data poisoning attacks against federated learning systems. In *European Symposium on Research in Computer Security*, pages 480–501. Springer, 2020.

Siyu Wang, Fangfang Liu, and Hailun Xia. Content-based vehicle selection and resource allocation for federated learning in iov. In *2021 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, pages 1–7. IEEE, 2021.

Chen Wu, Xian Yang, Sencun Zhu, and Prasenjit Mitra. Mitigating backdoor attacks in federated learning. *arXiv preprint arXiv:2011.01767*, 2020.

Chenming Xu and Yunlong Mao. An improved traffic congestion monitoring system based on federated learning. *Information*, 11(7):365, 2020.

Waleed Yamany, Nour Moustafa, and Benjamin Turnbull. Oqfl: An optimized quantum-based federated learning framework for defending against adversarial attacks in intelligent transportation systems. *IEEE Transactions on Intelligent Transportation Systems*, 2021.

Jiale Zhang, Junjun Chen, Di Wu, Bing Chen, and Shui Yu. Poisoning attack in federated learning using generative adversarial nets. In *2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data*

Science And Engineering (TrustCom/BigDataSE), pages 374–380, 2019. doi: 10.1109/TrustCom/BigDataSE.2019.00057.

Chen Zhao, Yu Wen, Shuailou Li, Fucheng Liu, and Dan Meng. Federatedreverse: A detection and defense method against backdoor attacks in federated learning. In *Proceedings of the 2021 ACM Workshop on Information Hiding and Multimedia Security*, pages 51–62, 2021.

Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. *Advances in Neural Information Processing Systems*, 32, 2019.

Bisher in der FAT-Schriftenreihe erschienen (ab 2019)

Nr.	Titel
316	MULTIC-Tooling, 2019
317	EPHoS: Evaluation of Programming - Models for Heterogeneous Systems, 2019
318	Air Quality Modelling on the Contribution of Brake Wear Emissions to Particulate Matter Concentrations Using a High-Resolution Brake Use Inventory, 2019
319	Dehnratenabhängiges Verformungs- und Versagensverhalten von dünnen Blechen unter Scherbelastung, 2019
320	Bionischer LAM-Stahlleichtbau für den Automobilbau – BioLAS, 2019
321	Wirkung von Systemen der aktiven, passiven und integralen Sicherheit bei Straßenverkehrsunfällen mit schweren Güterkraftfahrzeugen, 2019
322	Unfallvermeidung durch Reibwertprognosen - Umsetzung und Anwendung, 2019
323	Transitionen bei Level-3-Automation: Einfluss der Verkehrsumgebung auf die Bewältigungsleistung des Fahrers während Realfahrten, 2019
324	Methodische Aspekte und aktuelle inhaltliche Schwerpunkte bei der Konzeption experimenteller Studien zum hochautomatisierten Fahren, 2020
325	Der Einfluss von Wärmeverlusten auf den Rollwiderstand von Reifen, 2020
326	Lebensdauerberechnung hybrider Verbindungen, 2020
327	Entwicklung der Verletzungsschwere bei Verkehrsunfällen in Deutschland im Kontext verschiedener AIS-Revisionen, 2020
328	Entwicklung einer Methodik zur Korrektur von EES-Werten, 2020
329	Untersuchung zu den Einsatzmöglichkeiten der Graphen- und Heuristikbasierten Topologieoptimierung zur Entwicklung von 3D-Rahmenstrukturen in Crashlastfällen, 2020
330	Analyse der Einflussfaktoren auf die Abweichung zwischen CFD und Fahrversuch bei der Bestimmung des Luftwiderstands von Nutzfahrzeugen, 2020
331	Effiziente Charakterisierung und Modellierung des anisotropen Versagensverhaltens von LFT für Crashsimulation, 2020
332	Charakterisierung und Modellierung des Versagensverhaltens von Komponenten aus duktilem Gusseisen für die Crashsimulation, 2020
333	Charakterisierung und Meta-Modellierung von ungleichartigen Punktschweißverbindungen für die Crashsimulation, 2020
334	Simulationsgestützte Analyse und Bewertung der Fehlertoleranz von Kfz-Bordnetzen, 2020
335	Absicherung des autonomen Fahrens gegen EMV-bedingte Fehlfunktion, 2020
336	Auswirkung von instationären Anströmeffekten auf die Fahrzeugaerodynamik, 2020
337	Analyse von neuen Zell-Technologien und deren Auswirkungen auf das Gesamtsystem Batteriepack, 2020
338	Modellierung der Einflüsse von Mikrodefekten auf das Versagensverhalten von Al-Druckguss-komponenten mit stochastischem Aspekt für die Crashsimulation, 2020

339 Stochastisches Bruchverhalten von Glas, 2020

340 Schnelle, breitbandige Datenübertragung zwischen Truck und Trailer als Voraussetzung für das hochautomatisierte Fahren von Lastzügen, 2021

341 Wasserstoffkompatibilität von Aluminium-Legierungen für Brennstoffzellenfahrzeuge, 2021

342 Anforderungen an eine elektrische Lade- und Wasserstoffinfrastruktur für gewerbliche Nutzfahrzeuge mit dem Zeithorizont 2030, 2021

343 Objective assessment of database quality for use in the automotive research and development process, 2021

344 Review of non-exhaust particle emissions from road vehicles, 2021

345 Ganzheitliche Betrachtung von Rollwiderstandsverlusten an einem schweren Sattelzug unter realen Umgebungsbedingungen, 2021

346 Studie zur Abschätzung der Anwendungspotentiale, Risiken und notwendigen Forschungsbedarfe bei der Verwendung von Glashohlkugeln in Kombination mit thermoplastischem Schaumspritzguss, 2021

347 Typgenehmigungsanforderungen an Level-3-Autobahnsysteme - Hintergrundbetrachtungen zu technischen Anforderungen für eine automatisierte Fahrfunktion, 2021

348 Einfluss der Kantenbearbeitung von Aluminiumblechen auf das Restumformvermögen sowie die Festigkeitseigenschaften unter quasistatischer und schwingender Beanspruchung, 2021

349 Verstärkung dünner formgehärteter Bauteile mittels FVK-Verrippungen, 2021

350 HMI Anforderungen für den automatisierten Individualverkehr unter Berücksichtigung von Leistungsmöglichkeiten und -grenzen älterer Nutzer, 2021

351 Compatibility of polymers for fuel cell automobiles, 2021

352 Entwicklung einer gewichtsoptimierten Batteriegehäusestruktur für Volumenfahrzeuge, 2021

353 Charakterisierung und Modellierung des Deformations- und Versagensverhaltens von nicht-faserverstärkten Thermoplasten unter mehrachsiger Crashbelastung, 2021

354 Untersuchung zum thermischen Komfort im Pkw für den Grenzbereich des Luftzugempfindens, 2021

355 Anforderungen an die Güte, Verfügbarkeit und Vorausschau einer Reibwertschätzung aus Funktionssicht, 2021

356 Entwicklung einer standardisierten Prüfanordnung zur Bewertung der Übernahmeleistung beim automatisierten Fahren, 2022

357 Vorstudie zu Verkehrsemissionen - Räumlich und zeitlich aufgelöste Daten durch Schwarmmessungen, 2022

358 Produktivitätssteigerung und Kostensenkung der laser-additiven Fertigung für den Automobilbau, 2022

359 Analyse der Einflussfaktoren auf die Abweichung zwischen CFD und Fahrversuch bei der Bestimmung des Luftwiderstands von Nutzfahrzeugen mit Fokus auf den Ventilationswiderstand von Nfz-Rädern, 2022

360 Werkstoffmodelle und Kennwertermittlung für die industrielle Anwendung der Umform- und Crash-Simulation unter Berücksichtigung der thermischen Behandlungen beim Lackieren im Prozess bei hochfesten Werkstoffen, 2022

361 Compatibility of polymers for fuel cell automobiles, 2022

362 Ermüdung kurzfaserverstärkter thermoplastischer Polymerkwerkstoffe, 2022

363 Market research and definition of procedure to comparison of comfort measuring systems for a vehicle cabin, 2022

364 Methodische Ansätze zur Auswahl von Bordnetzstrukturen mit erhöhten Zuverlässigkeitsanforderungen, 2022

365 Fahrwiderstand von Lenk- und Liftachsen in Kurven und auf gerader Strecke unter realen Umgebungsbedingungen, 2022

366 Klimadaten und Nutzungsverhalten zu Auslegung, Versuch und Simulation an Kraftfahrzeug-Kälte-/ Heizanlagen, 2022

367 Experimentelle und numerische Untersuchung des selbsttätigen Losdrehens von Schraubenverbindungen mit konstanten und variablen Amplituden und Entwicklung einer Bewertungsmethode, 2022

368 Objective assessment of database quality for use in the automotive research and development process – Part 2, 2023

369 Level 2 hands-off – Recommendations and guidance, 2023

370 Funktionale Sicherheitsbewertung und Cybersecurity Analysen relevanter Use Cases für die Datenübertragung zwischen Truck und Trailer als Voraussetzungfür das hochautomatisierte Fahren von Lastzügen, 2023

371 Study on the technical evaluation of decentralization based de-identification procedures for personal data in the automotive sector, 2023

Impressum

Herausgeber	FAT Forschungsvereinigung Automobiltechnik e.V. Behrenstraße 35 10117 Berlin Telefon +49 30 897842-0 Fax +49 30 897842-600 www.vda-fat.de
ISSN	2192-7863
Copyright	Forschungsvereinigung Automobiltechnik e.V. (FAT) 2023

Verband der Automobilindustrie e.V. (VDA)
Behrenstraße 35, 10117 Berlin
www.vda.de
Twitter @VDA_online

Forschungsvereinigung Automobiltechnik e.V. (FAT)
Behrenstraße 35, 10117 Berlin
www.vda.de/fat

VDA Verband der Automobilindustrie

FAT Forschungsvereinigung Automobiltechnik