Defensive Publications Series

June 2020

DRIVER RECOGNITION AND DATA RETENTION BASED ON DRIVING BEHAVIOR

Verena Blunder
Bertrandt Ingenieurbüro GmbH

DRIVER RECOGNITION AND DATA RETENTION BASED ON DRIVING BEHAVIOR

Technical task:

The idea is to connect automated driver recognition by the use of machine learning algorithms within the car that can recognize a driver by its personal driving behavior with in-car data retention relevant for GDPR data privacy satisfaction.

Initial situation:

In the past, the driver of a car was not recognized by any automated/aided system at all. With personalization systems moving to the market, drivers are being recognized with the help of personal unique identifiers; these can reach from less complex system using a user/password combination to more complex systems connecting to a person's smartphone or equivalent products.

Next to this, General Data Protection Regulation (GDPR) requires a very high standard in personal data processing. As companies start to collect more data, it has to be ensured that customers either agree that data can be collected (with a certain reason) or disagree and no data is collected. Logically, this required the aforementioned recognition of driver/addressable person.

Referring to the aforementioned driver recognition procedures, it should be mentioned that the existing systems could be tricked if a driver's authentication method gets used by another person (e.g. stolen digital authentication or smartphone) or a user's authentication method is used although he's not the actual driver (e.g. it is not checked if the user is actually sitting on the driver seat and if so, the system is limited to the smartphone position in the vehicle).

Bringing both situations together, it can be deduced that companies could be hold accountable for data privacy breaches if wrong authentication leads to data collection of a person that has not approved the same.

Solution:

To explain the complexity of this new technical procedure, all parts will be explained separately:

- **Machine learning algorithms for driving behavior recognition**: there will be specialized algorithms used (e.g. deep learning algorithms) to monitor and recognize driving behavior. By an aggregation of multiple journeys, the technical solution could generate a digital image of individual driving behavior that matches a person to a high extent. This digital driving behavior image can then be bound to the already used authentication method (e.g. digital user).
- **Driver recognition**: If someone starts to drive with the car, a set of comparison algorithms is used to compare the digital image of the user to the actual driving behavior of the recent session. Naturally, this system has boundaries as it requires at least the generation of a digital image in advance and a certain (to be defined) time to drive so that the comparison algorithms can successfully work.
- **In-car data retention**: the whole set of data that is subject to GDPR data privacy regulations should be buffered within the car until there is sufficient evidence that the actual driver is matching the "user" that has approved data collection. At this point, data transmission can be started/continued. If there is not enough evidence or counterevidence, the data of this ride will be deleted, buffered for later approval or send as anonymized data package (depending on the main user's settings). Data from unauthorized rides could thereby be used for additional functions (see last dash within this list).
- **Evidence building**: if there is enough evidence should depend on multiple factors.
    - If there is no user logged in to the car, the algorithms function as a kind of user recognition comparing the recent driving behavior to the saved profiles of this specific car. As soon as there is significant evidence (extend has to be defined) for a match of actual driving behavior and digital image, the HMI should provide feedback to the driver that this user is now logged in (with the possibility to disagree) and data is getting collected or not collected according to the user's preferences.

    - If there is a user logged in to the car, the algorithms function as a kind of re-assessment of the recent user and should verify if the recent driving behavior matches the digital image of the user. If there is significant counterevidence (extend has to be defined) between user and driver of the car, the HMI should provide feedback to the driver that there is an approval for data collection missing or the authenticated user is probably wrong.
    - Potential further features like face/voice recognition or equivalent could be used as additional input criterion to deepen the evidence formula for user-driver matching.
- **Feature influencing**: additional features can be steered based on the result of the driver recognition process
    - Successful driver authentication could be linked to product activation (e.g. for function on demand or Connect products) and vice versa for unsuccessful authentication.
    - Unsuccessful authentication could be used for additional services, such as general anonymized data monetarization, owner's notification or vehicle tracking system activation.
    - Authentication checks in general could be linked to an authorization process (e.g. for sharing mobility purposes or car rental) or the car owner/main user.
    - By the introduction of the new procedure in general and specifically these functions, additional processes have to be provided to the owner/main user to deal with unauthorized accesses/rides to depict a legitimate digital image of the car's usage.

Advantages:

By the use of this technical solution, the driver recognition gets simplified and the security for the car manufacturer/dealer to comply with GDPR (and potentially further regulations) rises significantly.

Therefore, the solution solves two problems at a time: increasing customer satisfaction/usability and satisfaction of regulatory requirements, both using a technical verification for driver recognition and connecting this with the way data is retained in the car.

Thereby, human failure (e.g. if you forget to login to the car) and technical circumvention (e.g. if you try to trick the login by using stolen credentials) are excluded from this process. This can be also very valuable for the future, when shared driving becomes more usual and driver change happens so often that the driver recognition process is crucial for companies, both for offering competitive usability and having reasonable fraud prevention.

Possible application:

On a high-performance computer (HPC) within the car, the necessary algorithms for driving behavior recognition should be installed. These algorithms need to be detailed further. However, it is clear that they have to have the ability to create a digital image of driving behavior within minutes of driving with a certainty of more than 50% to recognize the actual driver. The comparison of actual driver vs. digital images should be working within seconds (up to 30 seconds max.) with a certainty of more than 80%.

Thereby, the initialization process is quite short and the distraction of wrong results by the procedure is reduced to a minimum from a customer perspective.

The technical implementation will be built as follows:

- Driving behavior algorithms (machine learning engine), driver recognition comparison algorithm and evidence procedures should run on a HPC within the car
- The user related digital images (of driving behavior) should be stored in a backend system to have them accessible from multiple cars
- The data retention service should run on a storage focused computer within the car and should be reactive; meaning that this service will basically store all data that is created during a ride based on the parameter input it gets from other HPCs (no need for high computational power); upon request from the HPC it sends out these data to a data lake for further

processing (in case no such request is placed, the data is buffered in an intelligent database on that computer)

- The HPC running the algorithms is responsible for running the algorithms, decision-making and in-car communication of necessary commands. As explained before, if the evidence engine recognizes sufficient evidence for a user-driver match, the HPC requests from a backend which approvals (for which data) this user has given and sends out a command to the storage computer that these data should be sent to the appropriate data lake.

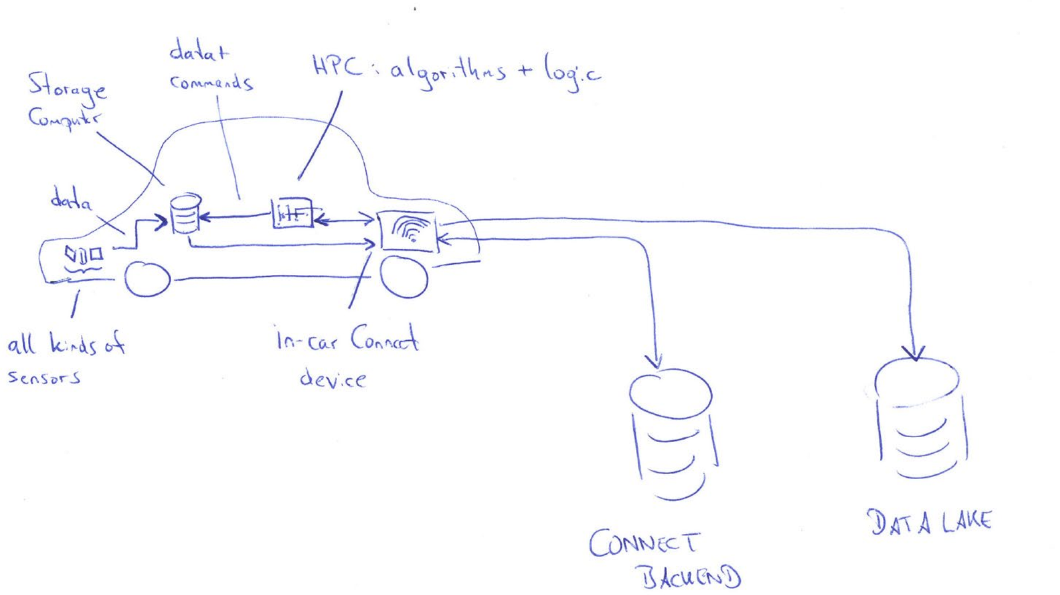

Figure 1: Rough technical architecture draft