16th ITS European Congress, Seville, Spain, 19-21 May 2025

Re-identification risks and privacy challenges in connected vehicle systems

Sven Maerivoet[1*], Bart Ons[1]

1. Transport & Mobility Leuven, Belgium

(*) sven.maerivoet@tmleuven.be

Abstract

The Trusted Integrity and Authenticity for Road Applications (TIARA) project aims to equip National Road Authorities (NRAs) with the tools and insights needed to develop secure and privacy-preserving data infrastructures for Cooperative Intelligent Transport Systems (C-ITS). In this paper we examine the privacy risks associated with C-ITS data, highlighting vulnerabilities in anonymisation methods and re-identification risks due to detailed data transmission, such as location and behavioural patterns. We explore state-of-the-art mitigation measures, including differential privacy, homomorphic encryption, secure multi-party computation, and federated learning, while emphasising the importance of frameworks like pseudonymisation, data minimisation, and transparent data sharing. Additionally, we discuss regulatory guidelines, such as the GDPR and ITS directives, which form the backbone of privacy and interoperability standards for connected vehicles in Europe. Our findings reveal significant privacy risks in C-ITS data, including vulnerabilities in anonymisation, re-identification threats through detailed data transmission, and the potential for behavioural profiling and location tracking, emphasising the need for advanced mitigation measures, robust regulatory frameworks, and collaborative efforts among stakeholders to ensure secure and privacy-preserving use of connected vehicle data. Future work will refine mitigation techniques, propose advanced measures, and foster stakeholder collaboration to ensure the secure, ethical, and privacy-preserving use of C-ITS data for traffic management and mobility innovation.

Keywords:

Data privacy, Connected vehicles

Introduction

Background of the TIARA project

The TIARA project (*Trusted Integrity and Authenticity for Road Applications*) provides National Road Authorities (NRAs) with an increased understanding of what is required to achieve a trustworthy and secure data infrastructure. The availability of data has allowed road users and NRAs to benefit from new business models. To deliver these benefits, the data infrastructure must be trustworthy and trusted, i.e. secure, with assurances that it is managed to achieve privacy for all stakeholders. As more C-ITS services develop in Europe, and road users access and share more C-ITS data through open border countries, NRAs will need to ensure greater interoperability through common approaches to connected systems. Data trust is therefore paramount.

CEDR is undertaking a series of projects to research how NRAs can maintain and share the digital road infrastructure data and improve the use of third-party data by NRAs. Since the C-Roads Platform has started, several ITS (*intelligent transportation systems*) programmes have been rolled out and it has been identified that there are key elements that the NRAs will need to understand before implementing these systems more widely. The TIARA project has been designed to address the two key areas of Trust and Privacy in C-ITS applications. The first subject Trust concerns an understanding of the implementation of trust models that could protect C-ITS data. The second subject Privacy concerns an understanding of the impact of processed user personal data, more specifically, the privacy impact of the processed road user location data, and recommendations to improve the location privacy-preservation for NRAs [1].

Terminology and nomenclature

In light of the different concepts that are related to data privacy protection, we hereby provide some explanatory terminology and nomenclature:

- **Anonymisation** is the process of removing or modifying personal information from a dataset so that individuals cannot be identified directly or indirectly. Once data is anonymised, it should be impossible to trace back to the original individual, making the data no longer subject to data protection laws (cf. GDPR).
- **De-anonymisation** is the process of reversing anonymisation. It involves re-identifying individuals from anonymised data by using additional information or advanced techniques. This process poses significant privacy risks, as it can potentially reveal sensitive information about individuals.
- **Pseudonymisation** is a data de-identification technique where personal identifiers are replaced with pseudonyms or artificial identifiers. Unlike anonymisation, pseudonymised data can be re-identified with the use of additional information kept separately.
- **Identification** refers to the process of recognising an individual within a dataset. It involves matching data points to a specific person, allowing the data to be attributed to that individual.
- **De-identification** is similar to anonymisation but generally refers to techniques that remove or obscure personal identifiers. Unlike anonymisation, de-identification does not always ensure that re-identification is impossible. De-identified data may still pose some risk of being re-identified under certain circumstances.
- **Re-identification** is the process of matching anonymised or de-identified data back to the individual it pertains to.

Anonymisation is a stronger form of data alteration, aiming to remove all identifiable information so that re-identification is not possible. De-identification, while similar, often leaves open the possibility of re-identification if additional information is available. Anonymised data is usually exempt from data protection laws, while de-identified data may still be subject to them depending on the risk of re-identification. Identification involves recognising an individual in a dataset for the first time, while re-identification involves matching de-identified or anonymised data back to the individual after the fact.

State of the art of mitigation measures

Anonymisation, access control, and data minimisation

Reducing the collection of personal data to the minimum necessary is a vital step in mitigating privacy risks, particularly in the context of connected vehicles. While pseudonymisation replaces direct identifiers with pseudonyms, its effectiveness is limited if the original dataset remains accessible, as indirect data can still lead to re-identification. Advanced methods, such as decentralised identity solutions and blockchain technology, can further enhance privacy by decentralising data control and ensuring secure, tamper-proof transactions. To bolster privacy, frameworks like Public Key Infrastructure (PKI) ensure secure communication in ITS. Dynamic consent models allow users to adjust their data-sharing preferences in real-time, offering granular control and compliance with privacy regulations. Complementary approaches, such as data spaces and role-based access controls, facilitate secure data exchanges while upholding privacy standards. Additionally, data reduction techniques like video coding and feature extraction retain utility while minimising privacy risks. Transparency in data collection and usage is essential for fostering trust.

Differential privacy and synthetic data

Differential privacy is a robust technique for enhancing data privacy by introducing controlled randomness, or noise, into data or the functions processing it. This approach obscures individual information even when aggregated datasets are shared or analysed, making it harder to trace data back to specific individuals. Techniques like the Laplace and Gaussian distributions are used to add noise, balancing privacy and data utility. However, excessive noise can reduce the accuracy and usefulness of the data, necessitating careful calibration. Applications of differential privacy include aggregating vehicle location data for traffic analysis, which protects individual locations while retaining aggregate insights, and analysing driving behaviour without compromising individual privacy. Synthetic data generation and geo-obfuscation also aim to preserve privacy while maintaining utility. Synthetic data structurally mimics real-world datasets but often struggles to capture the complexities of connected vehicle systems. Geo-obfuscation introduces imprecision in location data by adding noise or generalising to broader geographic areas.

(Homomorphic) encryption

End-to-end encryption ensures data remains encrypted from source to recipient, preventing interception and unauthorised access. Symmetric encryption, using a single key for both encryption and decryption, and asymmetric encryption, employing a pair of public and private keys, offer complementary approaches to data security. These methods protect data integrity and trustworthiness, supporting secure communications between vehicles and external systems like manufacturers and service providers. Homomorphic encryption takes data security a step further by allowing computations on encrypted data without revealing the raw data. This enables secure analysis of vehicular data by third parties, such as traffic management systems or insurers, while

preserving individual privacy. However, homomorphic encryption faces challenges, including significant computational overhead and limited support for complex operations, which may hinder its practicality for organisations like NRAs.

Secure multi-party computation

Secure multi-party computation (MPC) is a cryptographic technique that allows multiple parties to collaboratively compute a function over their inputs while ensuring that each party's individual data remains private. By leveraging complex protocols, MPC enables participants to contribute data, such as speeds or locations, to compute collective results — like average traffic speed — without exposing their individual inputs. This makes MPC particularly valuable in scenarios where raw data is too sensitive to share, providing both privacy protection and utility in deriving insights. For example, vehicles can securely contribute data to optimise traffic efficiency or safety measures without compromising personal data. MPC extends its utility beyond traffic management, enabling cooperative interactions between vehicles and urban infrastructure while preserving privacy. Vehicles can securely share data with traffic lights or parking systems to optimise routing and resource allocation, ensuring individual user data remains confidential.

Zero-knowledge proofs

Zero-knowledge proofs (ZKPs) are cryptographic protocols that allow one party to prove the truth of a statement to another without revealing any additional information beyond the validity of the statement. This ensures that sensitive data remains private during authentication or verification processes. In vehicle-to-anything (V2X) communications, ZKPs can be used to verify conditions such as a vehicle's priority at an intersection or eligibility for a restricted traffic lane without exposing detailed information like the vehicle's location or the identities of its occupants. ZKPs also have applications in financial transactions and subscription-based services in the automotive sector. For example, a vehicle can prove its active subscription to an automated toll system without revealing account details or personal information, streamlining the process while enhancing security.

Federated learning

Federated learning is a decentralised machine learning approach that enables multiple participants to collaboratively develop a shared model without sharing their raw data. Instead, the learning algorithm is trained locally on each participant's device, and only the updated model parameters or improvements are sent back to a central server. This ensures that sensitive data, such as driver behaviours, remains on the participant's device, significantly reducing the risk of data breaches and exposure. This approach not only prioritises privacy but also improves efficiency and scalability by eliminating the need for large centralised data storage and reducing network dependency. In the automotive sector, federated learning allows connected vehicles to collectively enhance safety features, fuel consumption, and contribute to traffic management strategies without exposing individual data.

Legal frameworks and regulations

In Europe, the privacy of connected vehicle data is governed primarily by the GDPR, supplemented by tailored guidelines from the European Data Protection Board (EDPB). These guidelines, adopted in 2021, emphasise principles like data minimisation, transparency, and robust security measures to ensure compliance with GDPR. They address various data categories, including location and biometric data, while detailing user rights to access, rectify, and delete their personal information. Complementary regulations such as the ITS Directive 2010/40/EU, its updates, and the draft C-ITS Delegated Act aim to ensure interoperability and secure data communication between vehicles and infrastructure. These frameworks foster the safe deployment of Cooperative Intelligent Transport Systems (C-ITS) and emphasise integrating security protocols like ISO-27001 to manage risks and safeguard user data. Beyond Europe, data privacy regulations for connected vehicles vary. In the United States, state laws and federal guidelines prioritise transparency and data minimisation, while China's Data Security Law and Japan's privacy regulations focus on protecting personal data in connected vehicles. OEMs view stringent privacy regulations like GDPR as critical for consumer trust but highlight challenges in balancing compliance, innovation, and cost. Manufacturers advocate for regulatory clarity and interoperability to avoid market fragmentation and enable sufficient data collection for vehicle safety and performance. Globally, achieving a balanced approach between privacy protection and technological advancement remains a key challenge for the connected vehicle ecosystem.

Challenges in data privacy

In a nutshell, the previous sections have shown that there are several pressing and recurring matters to consider when discussing mitigation measures. In summary we provide a concise overview of the most relevant aspects in this respect [5]:

- **Lack of transparency**: the complexity of data processing in connected cars makes it difficult to inform users clearly about what data is collected, by whom, and for what purpose. This complexity challenges the enforcement of privacy policies and user consent protocols.
- **Excessive data collection**: there is a risk that the vast amount of sensors and data collection points in connected cars could lead to unnecessary collection of personal data, not strictly required for the provided services.
- **Data retention**: proper data retention policies are crucial as there is a risk that data could be stored longer than necessary, increasing the risk of misuse or unauthorised access.
- **Control over personal data**: users often lack sufficient controls to manage their personal data effectively within connected car systems, which complicates the ability to maintain privacy.
- **Purpose limitation**: data collected for specific purposes, like vehicle maintenance, could be repurposed for other uses such as insurance adjustments or law enforcement surveillance without clear user consent.
- **Security risks**: as part of IoT, connected cars are susceptible to various security risks including cyberattacks, which could compromise both personal data and vehicle operation.

Impact study

Following our previously obtained insights, we now focus our attention on the next two points:

(1) What (types of) information about road users could be leaked from C-ITS data?

(2) What is consequently the potential impact on the data subject itself?

For (1) we first look at the specific content that is transmitted in the C-ITS data (i.e. what are the headers that are required per message type? which ones are optional? what message types are sent? what is the frequency with which the information is sent? etc.). Based on the information provided and their key features, we can check for possible links with personal data. We should not dismiss any attributes or characteristics, even if they seem unlikely. This is because what seems hard or unlikely now might not be in the near future. As local processing power increases, re-identification could become an issue, even if it doesn't seem urgent now.

Closely related to the previously described work, we will then investigate what and how great the impact of such data breaches are on a re-identified individual. As it may be possible to detect behavioural patterns, implicit sensitive information, or even other properties of such individuals, this research may also provide us with an idea of which part of the road users that can be re-identified on the basis of leaked location data, and what the required efforts are to accomplish this [6].

Information that can be leaked from C-ITS data

V2X (C-ITS) message types (sets) and their subtypes are closely related to the different types of services that are foreseen. In Table 1 we present some of the relevant message sets for V2X communications, currently already standardised or in the process of being standardised [2, 3, 4].

Table 1 – Relevant V2X communication message sets

Message	Description	Frequency	Main contents
BSM	Basic Safety Message	10 Hz (typically fixed)	Provides vehicle location, speed, heading, and other critical information for collision avoidance and traffic management
CAM	Cooperative Awareness Message	1-10 Hz	Transmits status information about a vehicle or road user to nearby vehicles and infrastructure
CPM	Collective Perception Message	1-10 Hz	Shares detected object information from a vehicle or infrastructure sensor systems to other vehicles and infrastructure
DENM	Decentralised Environmental Notification Message	Event-driven	Alerts nearby vehicles and infrastructure to hazardous events or conditions, e.g., weather, traffic jams, road works, etc.
IVIM	In-Vehicle Information Message	Event-driven or periodic	Provides in-vehicle signage and information, e.g., speed limits and warnings

MAP	Map Message	1 Hz or lower	Detailed road and intersection layout (cf. road and lane topology service and traffic light manoeuvre service)
MAPEM	Map Data Extended Message		Extends the MAP messages with additional information relevant to specific use cases
MCDM	Multimedia Content Dissemination Message	Event-driven or periodic	Distributes multimedia content to nearby vehicles or infrastructure
MCM	Manoeuvre Coordination Message	10 Hz	Coordinates manoeuvres between vehicles, such as lane changes or merging
PSM	Personal Safety Message	1-10 Hz	Broadcasts information about vulnerable road users (e.g., pedestrians, cyclists) to nearby vehicles and infrastructure
SPAT	Signal Phase and Timing Message	1-10 Hz	Provides information about the current and future status of traffic signals
SPATEM	Signal Phase and Timing Extended Message		Extends the SPAT messages with additional information relevant to specific use cases
SREM	Signal Request Extended Message	Event-driven	Allows vehicles to request signal priority or pre-emption (cf. traffic light control service)
SSEM	Signal request Status Extended Message	Event-driven	Provides status updates on signal requests, such as whether the request was granted

V2X messages typically contain a high-level structure that follows standardised formats to ensure interoperability. Any such logical grouping of related data elements within a message is referred to as a container. They help to organise the data in a structured and modular way, making it easier to interpret and process the information. Examples of these are:

(i) a **header**: message identification, protocol version, message length, etc.

(ii) a **payload**: the actual message content such as vehicle state data, etc.

(iii) **metadata**: e.g., timestamps, sender identification, geographical information, etc.

(iv) **security and integrity data**: digital signatures to ensure the authenticity and integrity of the message, along with extra encryption information

(v) **optional extensions**: e.g., custom data fields for application-specific data, error correction information, etc.

There exist common headers across all communication messages, i.e. **Message ID** (every message type includes a unique identifier to distinguish it from other messages), **Station ID** (most messages include a unique identifier for the sending or receiving station, e.g., vehicle, roadside unit, etc.), and while not universally present in every single message type, the **Generation Time** (to indicate the time when the message was generated, thereby providing a timestamp for the message) is quite common in many of them. These common headers ensure that each message can be uniquely identified, traced back to its source, and properly time-stamped for effective communication and processing.

Privacy-preserving measures

According to [2] there are a number of measures that have the goal of guaranteeing privacy in V2X messages and their communication: **Pseudonymisation** ensures that the data transmitted cannot be directly linked to a specific individual. This involves using pseudonyms that can only be related to an individual through the collusion of two certification authorities, and only if these authorities have archived the relevant information. **Controlled data elements** whereby the data elements in, e.g., CAMs are carefully selected to exclude any information that can directly identify a vehicle, its owner, or its driver. As mentioned before, data like license plates, registration information, VINs, and other so-called persistent identifiers are not included. This minimises the risk of personal identification from the transmitted messages. **Frequently changing identifiers** ensure that the continuous reception of V2X messages from the same vehicle does not allow for the reconstruction of a vehicle's journey. **Limited data retention** ensures the received messages are not retained longer than necessary. For example, driving conditions data are kept only for a few seconds to minutes, depending on the service's needs, and are erased once the emission conditions are over. **Data minimisation and frequency control**: The frequency of transmission is minimised to the bare essential where possible so as to balance privacy and safety. Typically, European standards reduce the transmission rate to the necessary minimum, considering the driving situation and vehicle speed. **Silent periods**: By introducing silent periods between certificate changes, we can mitigate the risk of tracking by making it more difficult to link consecutive messages to the same vehicle and thus reduce the likelihood of continuous tracking. **Non-relay of messages**: In order to prevent widespread tracking, received CAMs are not forwarded, nor multi-broadcasted. This restriction limits the data's reach, ensuring that only vehicles within immediate proximity can access the information. **Segregation of duties**: The system design incorporates segregation of duties among different authorities to control access to data. The linking of Authorisation Tickets (ATs) to Enrolment Certificates (ECs) is managed by the Authorisation Authority, while linking ECs to vehicle communication unit numbers is handled by the Enrolment Authority. This separation ensures that no single entity can track a vehicle without colluding with multiple authorities, thereby enhancing privacy protection.

Correlations with personal data and potential impact on data subjects

Given the concern that increased local processing power and advances in data analytics could make re-identification and privacy issues more pronounced in the future, we now focus on additional potential correlations and privacy risks associated with V2X data that have potential impacts on data subjects.

- **Behavioural biometrics and re-identification risks**: Driving styles, vehicle usage patterns, and location traces can act as unique biometric identifiers, posing significant re-identification risks, particularly when combined with external data sources like public records or social media.
- **Cross-referencing with other data sources and behavioural profiling**: Integrating C-ITS data with smart home devices, social media, and public records enables detailed behavioural profiling by revealing daily routines, driving habits, and lifestyle choices, increasing re-identification risks.
- **Temporal and spatial analysis with location-based inferences**: Long-term movement data analysis can reveal detailed travel patterns, predict future behaviours, and infer sensitive information such as home and work locations, affiliations, and personal interests.
- **Event participation and inferred social connections**: DENMs can reveal involvement in road incidents, impact insurance and liability, and infer social or professional relationships based on shared routes, frequent proximity, and travel patterns.
- **Economic and financial inferences with vulnerable road users**: Vehicle details, travel habits, and PSMs can reveal economic status, spending preferences, and patterns of vulnerable road users, raising privacy concerns and risks of profiling or exploitation.
- **Health and wellness indicators from predictive analytics**: Travel patterns to medical facilities, gyms, or recreational areas can reveal sensitive health conditions, activity levels, and fitness habits, offering detailed insights into an individual's health and wellness.
- **Enhanced personal threats**: Detailed movement and behavioural profiles can enable intrusively targeted advertising based on frequent routes or destinations, increasing the risk of stalking, theft, or even physical harm as criminal activities.

Conclusions and remaining work

Our paper provided a comprehensive analysis of re-identification risks and privacy challenges associated with anonymised data in connected vehicle systems. We highlighted the susceptibility of C-ITS messages, containing detailed location and behaviour data, to deanonymisation, even when basic anonymisation techniques are applied. Through literature reviews and case studies, We examined re-identification methods and trends, such as leveraging minimal data points or integrating deep learning, while also exploring mitigation measures like differential privacy, synthetic data generation, and encryption. Despite safeguards like pseudonymisation and frequently changing identifiers, C-ITS data still presents privacy concerns due to the sensitive information it transmits, which can reveal travel patterns, real-time tracking, and even social relationships. Our study also assessed the correlation between V2X data and personal information, showing how repeated location or temporal data can lead to detailed profiling and re-identification. While the benefits of C-ITS for traffic management and safety are significant, advancements in data analytics heighten privacy risks, necessitating

robust protection measures and collaborative approaches among stakeholders to ensure ethical and responsible use of re-identification technologies.

Our current and future work focuses on evaluating the effectiveness of current re-identification mitigation measures and proposing additional strategies to make re-identification reasonably difficult. This includes exploring advanced methods for obfuscating vehicle identifiers, assessing the impact of location and temporal data thresholds, and addressing specific risks for individuals with higher re-identification odds, such as commuters or pedestrians. Furthermore, we will analyse potential pitfalls in existing approaches and examine how evolving technologies might exacerbate privacy risks. Building on these insights, we will consolidate findings into actionable recommendations to NRAs, offering guidelines for handling new data use cases and ensuring robust privacy protection as data sharing in connected vehicle ecosystems becomes increasingly prevalent. Through our interactions with stakeholders, we will also emphasise the importance of an information classification framework to better assess privacy risks and impacts.

Acknowledgments

The TIARA project is funded in the CEDR 2022(2) Research call on Data. An experienced team of European research organisations have gathered under the coordination of AESIN/Techworkshub, the UK-based member trade association. SINTEF, as an independent and non-profit research organisation, has independent technical expertise and deep experience from PKI deployments in multiple sectors. Traficon has longstanding experience of independent work with NRAs, specifically legal and ethical expertise of particular relevance to this project. TML, bridging the gap between university and private sector, is an independent open and transparent organisation with extensive experience of data analyses and privacy ramifications.

References

1. Maerivoet, S., and Ons, B., (2024). *Draft report on connected vehicle deanonymisation research review and impact study*, CEDR TIARA project Deliverable D8, October 2024.
2. C2C-CC (2018). *FAQ Regarding Data Protection in C-ITS*, CAR 2 CAR Communication Consortium, 18 September 2018.
3. Rondinone, M. and Correa, A. (2018). *Definition of V2X Message Sets*. TransAID Horizon 2020 Deliverable D5.1.
4. C-Roads Platform (2023). *C-ITS Message Profiles*, Working Group 2 Technical Aspects, Taskforce 3 Infrastructure Communication, version 2.1.0, 14 December 2023.
5. Rebiger, S., Moraes, T., Lareo López de Vergara, X., and Zerdick, T. (2019). *Connected Cars*, EDPS Tech Dispatch, issue 3.
6. de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V.D. (2013). *Unique in the Crowd – The privacy bounds of human mobility*, Scientific Reports.