Privacy-Aware Vehicle Data Handling: Contextual Access Control, Policy-Based Enforcement, and Utility-Balancing Analytics

Abstract

Connected vehicles continuously generate large volumes of sensitive sensor data that are valuable for mobility services, insurers, and third-party applications. However, sharing such data introduces security and privacy risks. These include unauthorized tracking of drivers, long-term profiling of driving behavior, inference of sensitive attributes, and misuse or disclosure of personal data by third parties. Such risks must be mitigated in compliance with data protection regulations such as the GDPR. This dissertation addresses these challenges by proposing a privacy-aware data governance framework for connected vehicles that encompasses in-vehicle access control, cross-organizational data sharing, and privacy-preserving data analytics.

First, at the access control layer, the research identifies automotive-specific requirements, namely time-based, location-based, and frequency-based data access and proposes an extended XACML-based model called XACML4M (XACML for Mobility). XACML4M introduces components such as the Vehicle Data Environment, Time Extension, Geolocation, and Polling Frequency Providers to enable fine-grained, dynamic, and context-aware access control. A functional prototype was implemented on a Raspberry Pi 4 and validated through real-world use-case policies, confirming the feasibility of enforcing dynamic authorization in vehicular environments.

Second, for cross-organizational data sharing, the dissertation presents Sticky-PRE (Sticky-Proxy Re-Encryption), a protocol that combines machine-readable sticky policies with proxy re-encryption. Sticky-PRE ensures that user-defined access conditions persist throughout the data lifecycle, even across organizational boundaries, while maintaining data confidentiality, integrity, and accountability. The protocol was analyzed under an honest-but-curious threat model and evaluated quantitatively through encryption, re-encryption, and decryption performance benchmarks, demonstrating secure and efficient operation suitable for connected vehicle contexts.

Finally, for downstream data analytics, the research examines the privacy-utility trade-off in vehicle insurance analytics using real-world anonymized data. The study systematically evaluates how anonymization techniques, k-anonymity, and l-diversity (distinct, shannon entropy, and recursive (c,l)-diversity) affect machine-learning model performance for insurance claim prediction. The results show that it is possible to substantially reduce re-identification risk while maintaining competitive predictive performance. For example, a configuration based on Distinct l-diversity with l=10, combined with a support vector classifier, achieves a residual risk of 0.049% and a k-anonymity level of 2013 with only minor losses in prediction accuracy compared to the original dataset. These findings provide practical guidance on how organizations can select anonymization parameters that balance data protection and analytical value.

Collectively, these interconnected contributions realize a comprehensive framework for secure, accountable, and privacy-preserving vehicle data sharing that aligns with modern data-protection principles while enabling meaningful data-driven services.