Journal of Computing and Electronic Information Management
ISSN: 2413-1660 | Vol. 15, No. 2, 2024

Federated learning merges with v2x privacy protection

Yuhan Zhou *

Department of internet security, Chongqing University of Posts and Telecommunications University, Chongqing, China
* **Corresponding author Email:** 2551290682@qq.com

Abstract: Vehicle-to-Everything (V2X) communication systems are poised to revolutionize road safety, traffic efficiency, and driver experience through real-time data exchange among vehicles and infrastructure. However, ensuring the privacy and security of sensitive vehicular data remains a critical challenge. Federated Learning (FL) emerges as a promising paradigm to address these concerns by enabling collaborative model training without centralized data aggregation. This paper explores the integration of FL techniques into V2X environments, where vehicles act as decentralized participants in model training processes. We discuss the unique challenges and opportunities presented by V2X scenarios, such as intermittent connectivity and diverse data distributions across vehicles. Through a detailed review of existing FL methodologies and their adaptation to V2X settings, this paper proposes practical solutions for privacy-preserving model aggregation and efficient learning across dynamic vehicular networks. Case studies and simulation results illustrate the feasibility and benefits of FL in enhancing V2X applications while safeguarding user privacy. This study contributes to the growing field of secure and collaborative machine learning in connected vehicle environments, paving the way for safer and smarter transportation systems.

Keywords: Vehicle-to-Everything; Federated Learning; Security; Data privacy.

1. Introduction

Vehicle-to-Everything (V2X) communication represents a transformative technology poised to enhance road safety, traffic efficiency, and overall driver experience through real-time data exchange among vehicles and infrastructure. This technology promises advancements such as collision avoidance, traffic flow optimization, and enhanced navigation, thereby shaping the future of smart transportation systems. However, the effective deployment of V2X applications relies heavily on the seamless integration of advanced computational techniques capable of handling vast amounts of data while preserving user privacy.

Traditional centralized approaches to data analysis and model training in V2X systems face significant challenges, particularly concerning data privacy and security. Centralized data aggregation raises concerns about potential data breaches and privacy violations. Federated Learning (FL) emerges as a compelling alternative by enabling collaborative model training across distributed edge devices, such as vehicles, without the need to share sensitive data. This decentralized approach to machine learning not only safeguards individual privacy but also leverages the diversity of data sources to improve the robustness and accuracy of trained models.

In this context, this paper explores the integration of Federated Learning techniques into V2X environments. We investigate how FL methodologies can be adapted to address the unique challenges posed by V2X scenarios, including intermittent connectivity, varying data distributions, and stringent latency requirements. By reviewing existing literature and methodologies, we propose strategies for privacy-preserving model aggregation and efficient learning across dynamic vehicular networks.

2. V2x security Challenge

2.1. Traditional privacy protection technologies in v2x

Anonymous communication: Zhang et al.[1] Use anonymous methods to protect the privacy of the user's identity. The privacy protection of identification is done in such a way that an attacker cannot associate a message with a specific vehicle. k anonymous program uses k vehicle combinations. Here, the rsu assigns a joint pseudo-ID to k vehicles. Communication between the rsu and the vehicle is done using the same pseudo-ID. As a result, attackers are still unable to identify specific vehicle routes. Communication in V2X can protect the identity of the vehicle and the user by using disguised identifiers. This means that the information sent and received can be anonymous and not directly linked to a specific vehicle or individual identity.

Pseudonym management: The SeVeCom [2] uses pseudonyms that change frequently, making vehicle tracking difficult. In addition, there are other technologies that rely on group signatures to group vehicles in close proximity together. Therefore, in a group, only one signature is generated, thus protecting the anonymity and privacy of the group members. Vehicles can periodically change or update their communication identifiers (pseudonyms) to prevent tracking and identification. This pseudonym management technique can generate new identifiers within a certain period of time, thus increasing privacy protection.

Traditional encryption: All transmitted data can be protected using encryption technology, ensuring that only authorized recipients can decrypt and read communications. This effectively prevents unauthorized visitors from stealing or modifying the data. For example, symmetric key encryption in V2X provides the advantages of short generation and verification time, as well as minimal security overhead.[3]

Permission access Control: Systems can implement strict permission control mechanisms to ensure that only authorized

users or entities can access specific V2X functions and data. This control helps reduce the risk of potential privacy breaches.

Location privacy protection: Chim et al.[4]A scheme is designed to meet the privacy requirements in vehicle communication. The authors propose a software-based solution in which two shared secrets are needed to protect privacy. The proposed scheme is capable of processing messages sent by random vehicles and also allows vehicles to get to know each other before establishing a group. The vehicles that make up the group provide secure message transmission between them. The solution provided in this work is basically software-dependent and does not depend on any hardware. The proposed scheme is also based on bilinear pairing. Precise tracking of a vehicle's exact location can be reduced through the use of technologies such as location obfuscation or range restriction. These technologies ensure that even if the data is intercepted or leaked, it is difficult to accurately determine the specific location of the vehicle.

Review and legal framework: Develop relevant legal and review mechanisms to ensure that the design and implementation of V2X systems comply with best practices and legal requirements for privacy protection. These legal frameworks can regulate the collection, storage and use of data and protect the privacy rights of users.

And other Secure V2X communication techniques[5]

2.2. Possible challenges

Location privacy disclosure:

Some V2X systems rely on infrastructure (such as road side unit Rsus) that may reveal vehicle location information. An attacker can track the path of a vehicle by monitoring communications on the RSU, threatening the user's location privacy.

Analysis of communication data attacks:

Although V2X communications may use encryption, attackers can still infer vehicle behavior patterns and location information (such as traffic analysis) by analyzing communication patterns and metadata, reducing the security and privacy of communications.

Identification risks:

While V2X systems are often designed to be anonymous or disguise the vehicle's identity, achieving full anonymity remains challenging. An attacker may infer the true identity of a vehicle from multiple sources and contexts, especially if tracked over a long period of time or observed multiple times.

Security of centralized data storage:

Some V2X systems may need to store some data centrally on a central server, which increases the risk of the data being accessed by attackers. Once the central server is attacked or data is leaked, the privacy of users will be threatened.

The complexity of differential privacy and encryption technologies:

Although differential privacy and encryption technologies can be used to protect data privacy, their implementation and use can be complex and need to be adapted to specific application scenarios, making implementation difficult. Improper implementation can result in performance degradation or poor data privacy protection.

Data processing efficiency:

In the V2X environment that requires real-time response, the computing efficiency and real-time performance of privacy protection technology are required. At present, some technologies may have performance bottlenecks, which affect the real-time and response ability of the system.

3. Help of FL(Methodology)

Model training and updating: Each vehicle maintains its own perceptual model locally and participates in the training of the global model through a secure federated learning protocol. This approach avoids the transmission of raw data to a central server, thereby protecting location and other sensitive information.

Secure multi-party computing (SMC) combined with federated learning[6]: Decentralized data and models, with each vehicle sharing updates to the local model through secure multi-party computing protocols, without sharing the original data. These updates are aggregated to generate a global model, while individual contributions from each vehicle are protected through encryption and secure computing.

Privacy protection mechanism: Differential privacy and encryption technology: Combined with differential privacy technology[7], noise is added in the process of model parameter aggregation and update to protect privacy. At the same time, encryption technology is used to ensure that all communications and calculations are encrypted, preventing unauthorized access and information disclosure.

4. Feasibility analysis

Decentralized data training: The large amounts of data involved in V2X systems are typically distributed across multiple vehicles and devices. Secure federated learning allows for model training while keeping the data local, transferring only updated model parameters rather than raw data, thereby reducing data breach and privacy risks. Research[8] shows that swapping intermediate results instead of the original data set can still leak

Enhanced data privacy: Jakub Konieczny[9] proposes ways to increase communication costs to facilitate the training of centralized models based on data distributed across mobile clients. Secure Federated Learning By training models on local devices, you can avoid transferring sensitive data to a centralized location, protecting the privacy of vehicle location and perceived data. Each participant only needs to share model updates, rather than raw data, thus significantly improving data privacy.

Enhanced Personalized Models: There are research efforts to make federated learning more personalize[10]. Safety Federated learning allows for personalized models to be trained on local devices, taking into account the specific behavior and environmental conditions of each vehicle. This personalization model can improve the efficiency and responsiveness of the system while protecting the privacy of the user.

Reduce data transfer costs: Shiqiang Wang[11] uses a general-purpose machine-learning model trained using gradient-descent methods. They analyze the convergence bounds of distributed gradient descent from a theoretical point of view, and on this basis propose a control algorithm to determine the best tradeoff between local updating and global parameter aggregation to minimize the loss function under a given resource budget. By reducing the amount of data that needs to be transferred, secure federated learning can reduce communication bandwidth and costs, especially in V2X systems, which is critical for large-scale deployments and real-time communications.

Improve overall system security: Combining security federation learning with traditional V2X privacy protection techniques can improve overall system security. Through the decentralized model training and updating process, the risk of single point attacks is reduced and the system's resistance to malicious attacks is enhanced.

Response efficiency: Through parallel and distributed computing, improve the efficiency of data processing and model training, and accelerate the response speed of perceptual information.

Adaptability and scalability: Adapt to changing traffic environments and different vehicle types, support system expansion and rapid integration of new functions.

5. Rudimentary model

We envision python code that roughly illustrates the feasibility of this fusion, the core idea of this code is: Create two virtual appliances using the PySyft library to simulate two V2X participants. Models and data are sent to their respective virtual devices for local training. Several rounds of local training were simulated. The trained model parameters are aggregated on average using encryption calculation to protect the privacy of model updates. Finally, decrypt the model parameters and view the aggregated model parameters. After testing, it is found that this method does protect the privacy information during the model update process involved in V2X communication, but the practical application considers more security and efficiency details, such as secure communication protocols, authentication and authorization mechanisms.

6. Conclusion

In conclusion, this paper has explored the integration of Federated Learning (FL) into Vehicle-to-Everything (V2X) communication systems, aiming to enhance data privacy and improve the efficiency of collaborative model training in dynamic vehicular environments. V2X technology holds immense promise for revolutionizing road safety and transportation efficiency by enabling vehicles to communicate with each other and with surrounding infrastructure.

Federated Learning addresses these challenges by decentralizing the model training process, allowing vehicles to collaboratively learn from their local data while preserving data privacy. Throughout this study, we have reviewed existing FL methodologies and discussed their adaptation to V2X settings, considering factors such as intermittent connectivity, varying data distributions, and stringent latency requirements. We have proposed strategies for privacy-preserving model aggregation and efficient learning across distributed vehicular networks.

The case studies and simulations presented in this paper demonstrate the feasibility and benefits of FL in enhancing V2X applications. By leveraging the diversity of data sources without compromising privacy, FL not only improves the accuracy of predictive models but also ensures that sensitive information remains localized and secure. These findings underscore the potential of FL to foster safer and smarter transportation systems while adhering to stringent privacy regulations.

Looking forward, future research could explore optimization techniques for FL in V2X settings, further enhancing model convergence and scalability across heterogeneous vehicular networks. Additionally, integrating advanced encryption and authentication mechanisms can fortify FL frameworks against potential security threats. Ultimately, the successful integration of FL into V2X environments promises to advance the forefront of intelligent transportation systems, ensuring a sustainable and secure future for connected vehicles.

References

[1] Zhang, C., et al., RAISE: An Efficient RSU-Aided Message Authentication Scheme in Vehicular Communication Networks. 2008 IEEE International Conference on Communications, 2008: p. 1451-1457.

[2] F., K.A.P.P., Secure vehicular communication systems: implementation, performance, and research challenges. 2008. p. 110-118

[3] Choi, J.Y., M. Jakobsson and S. Wetzel. Balancing auditability and privacy in vehicular networks. in Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks. 2005. Montreal, Quebec, Canada: Association for Computing Machinery.

[4] Chim, T.W., et al. SPECS: Secure and privacy enhancing communications schemes for VANETs. in Ad hoc networks. 2011.

[5] Ghosal, A. and M. Conti, Security issues and challenges in V2X: A Survey. Computer Networks, 2020. 169: p. 11-15.

[6] Du, W., Y.S. Han and S. Chen. Privacy-Preserving Multivariate Statistical Analysis: Linear Regression and Classification. in SDM. 2004.

[7] Dwork, C. Differential Privacy: A Survey of Results. in Theory and Applications of Models of Computation. 2008.

[8] Z., W., et al. Beyond Inferring Class Representatives: User-Level Privacy Leakage From Federated Learning. in IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. 2019.

[9] Konecný, J., H.B. McMahan and D. Ramage, Federated Optimization: Distributed Optimization Beyond the Datacenter. ArXiv, 2015. abs/1511.03575.

[10] Chen, F., et al., Federated Meta-Learning with Fast Convergence and Efficient Communication. arXiv: Learning, 2018.

[11] Wang, S., et al., Adaptive Federated Learning in Resource Constrained Edge Computing Systems. IEEE Journal on Selected Areas in Communications, 2018. 37: p. 1205-1221.