Security and privacy in V2X communications: how collaborative learning can improve cybersecurity?

Pradip Kumar Sharma
Deepansu Vohra
Shailendra Rathore

Sharma, P.K., Vohra, D. & Rathore, S. (2022) 'Security and privacy in V2X communications: how collaborative learning can improve cybersecurity?'. *IEEE Network*, 36(3): pp.32-39. DOI: https://doi.org/10.1109/MNET.003.2100522

Security and Privacy in V2X Communications: How Collaborative Learning can Improve Cybersecurity?

Pradip Kumar Sharma, *Senior Member, IEEE*, **Deepansu Vohra**, **Shailendra Rathore**

***Abstract*—Advances in cellular technology are a key driver of the growing automotive Vehicle to Everything (V2X) market. In V2X communications, information from sensors and other sources travels via high-bandwidth, low-latency, high-reliability links, paving the way to fully autonomous driving and intelligent mobility. With the future adoption of 5G and beyond (5G&B) networks, V2X is likely to generate a huge volume of data, which encourages the use of edge computing and pushes the system to learn the model locally to support real-time applications. However, the edge computing paradigm raises concerns about the security and privacy of local nodes (e.g., vehicles) and the increased risk of cyberattacks. In this article, we identify open research questions, key requirements, and potential solutions to provide cyber resilience in V2X communications.**

***Index Terms*—V2X, Autonomous Vehicles, 5G&B Networks, Cybersecurity, Collaborative Learning**

I. Introduction

With the adoption of connected and autonomous vehicles, the automotive industry has gone beyond Vehicle to Everything (V2X) communications that encapsulates other communications mechanics such as Vehicle-to-Network (V2N), Vehicle-to-Infrastructure (V2I), and Vehicle-to-Vehicle (V2V). In V2X, information from sensors and other sources travels via high-bandwidth, low-latency, high-reliability links, and is seen to pave the way to fully autonomous driving. The development of Next-Generation Networks (NGNs) integrate the functionalities of emerging technologies such as data-plane programmability, Artificial Intelligence (AI), 5G&B, and Software-Defined Networking (SDN) to support V2X communications. NGNs enable significant advancements in imaging, presence technologies, and location awareness in V2X and it is to be much more heterogeneous than their predecessors, and are most likely to support intelligence services and applications beyond current mobile use scenarios. However, the increasing adoption of NGNs in V2X is being generating vast amount of data, which encourages hackers to threats the security and safety of connected and autonomous vehicles infrastructure. The hackers can manipulate data transmission that can weaken real-life safety. Also, successful attempts to attacks on communication and V2X infrastructure can impact all endpoints that potentially lead to danger of not only drivers lives but become highly vulnerable to so many more lives. For instance, if an attacker hacks a car then he/she can hack every car in the infrastructure because everything connected to each other autonomously from charging station to power plant that lead to cyber risks to national infrastructure. All the utilities of infrastructures such as traffic lights can be exploited by the attacker to mismanage the traffic of the city. Several automotive and V2X industries, including Bosch and Continental have already begun focusing on to drive the V2X cybersecurity market. The MARKETSANDMARKETS predicated 33.5% CAGR growth in the global V2X cybersecurity market size (i.e., from USD 659 million in 2020 to USD 2,798 million by 2025). Consequently, the objective our study to identify the importance of security and privacy in V2X infrastructure and build a zero trust and reputation cybersecurity model for autonomous vehicles in V2X communications. The main contributions of this article are as follows:

- We identified the key performance affecting cyber security to assess NGNs, particularly 5G&B networks in V2X infrastructure.
- The security and privacy aspects towards 5G&B are discussed and depicted in V2X environment.
- We design and experimentally analyze a security scenario of a data exchange model for autonomous vehicles in V2X communications.

II. Key Requirements of Network Architecture

The NGNs (i.e., 5G&B) will be the main catalyst for the realization of the autonomous ecosystem. In V2X communications, as the automotive industry moves towards fully autonomous driving, vehicles must communicate with each other and with road infrastructure. Dealing with the multitude of information necessary for safe driving is complex, and ultimately communication between human drivers must be fully reflected in the future [1-2]. In this section, we identify the key performance indicators for the assessment of 5G&B networks as follows:

Ultra-low latency: The network must be designed to provide ultra-low latency communication that is optimized to handle a

P. K. Sharma is with the Department of Computing Science, University of Aberdeen, UK (pradip.sharma@abdn.ac.uk)

D. Vohra is with the Department of Computing Science, University of Aberdeen, UK (d.vohra.20@abdn.ac.uk)

S. Rathore is with the Division of Cyber Security, School of Design and Informatics, Abertay University, UK (s520967@uad.ac.uk)

very large volume of data and to respond to rapidly changing data in real-time access. The intelligent services and application of V2X communication such as collaborative autonomous driving needs ultra-low latency to facilitate real-time control and management among vehicles and other network utilities such as traffic lights.

Localization and sensing: The network should be intelligent to enable high precision localization and sensing services for V2X communications. Context-aware intelligent networks will be able to leverage localization and sensing information to optimize deployment, operation and energy consumption without human intervention [3]. It helps deploying V2X network elements with a known position.

Distributed and in-network computation: How quickly the network adapts to new conditions is a critical factor in a self-sustaining ecosystem. Distributed and in-network computing is a key measure to achieve full automation without manual intervention in V2X communications.

Higher data rate: Considering V2X communications service scenarios, the network architecture should be designed to allow much faster sample rates, throughput, and data rates.

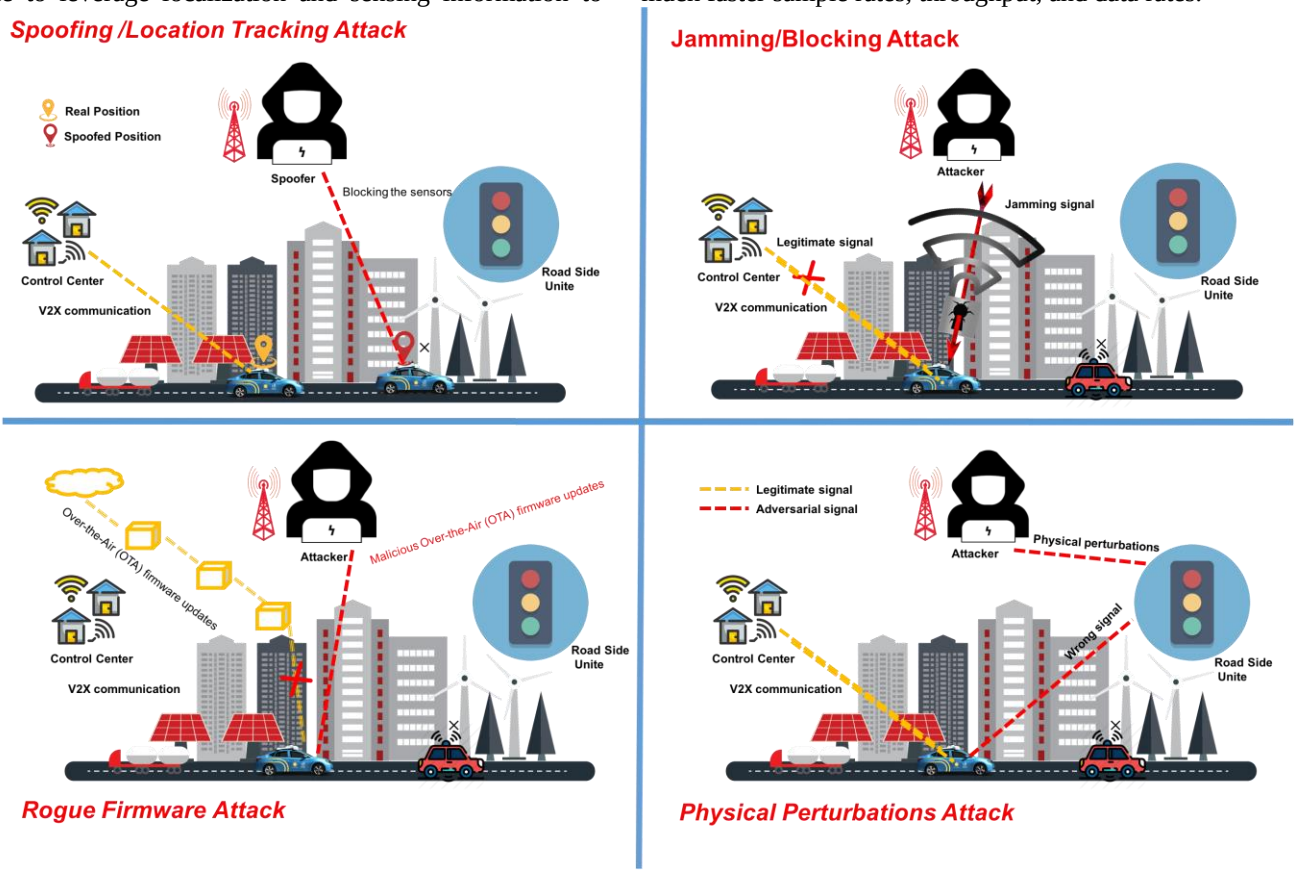


Fig. 1 Security attacks in V2X communications

Ultra-connectivity: Connectivity will be the critical measure in the realization of the autonomous ecosystem in 5G&B networks. Several sensors and cameras in V2X communications are responsible for knowing the context in the environment. Increased connectivity will help improve traffic flow and be an important step on the road to autonomous driving.

User-centric mobility: When it comes to the autonomous ecosystem, user preferences will shift more towards autonomous mobility. The development of autonomous vehicles in V2X communications should place more emphasis on individual mobility than on public mobility without human intervention.

Higher energy efficiency: Energy efficiency is essential to the long-term network service in V2X communications. The energy consumption of user terminals will also increase with the rapid growth of high data throughput. Considering constraints such as limited size, space and battery capacity, higher energy efficiency has become a key performance indicator in 5G&B networks.

III. Security and Privacy Issues

In this section, we discuss the impediment in V2X communications in terms of security and privacy aspects towards 5G&B.

A. Security

In 5G&B networks, V2X communications can be used to exchange information with other vehicles or traffic infrastructure. The SAE [4] standard defines six levels of driving automation for road vehicles, ranging from level 0 without driving automation at all to level 5 with full driving automation and without a driver. To enable automation from level 3 onwards, a higher level of computational functionality and connectivity is required, which, in turn widen the attack

surface and the likelihood of physical and cyber-attacks [5, 6]. As shown in Fig. 1, these attacks consist of the following:

Location tracking: Applications based on V2X communications rely more on the exchange of information with other vehicles or on the traffic infrastructure, which involves the sharing of the location with other local entities. Attackers can easily collect and misuse this information to track users.

Spoofing attack: The attacker can provides bogus information to the vehicles, users, and roadside infrastructures, which can lead to activities that are detrimental in V2X communications.

Physical perturbations attack: To perceive false information about the surrounding environment, adversaries can introduce

physical perturbations to traffic signs and road markings. The attacker can carefully create patterns such as the projection of light on the road sign and the lanes or the placement of stickers. Such a physical perturbations attack can cause the perception of wrong information by vehicles, users, and road infrastructure.

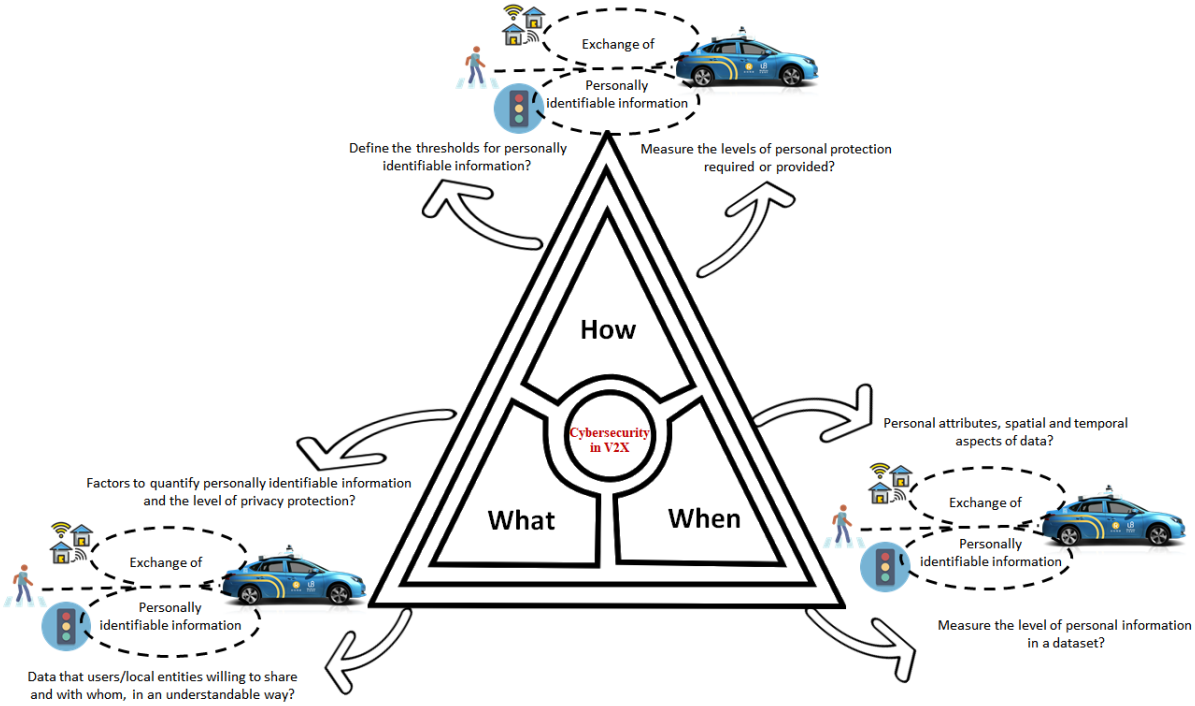


Fig. 2 Privacy issues in V2X communications

Rogue firmware attack: Installing rogue firmware on local entities is a critical concern for cybersecurity experts. An adversary can remotely discover an exploitable vulnerability and deploy malicious firmware from back-end servers. Malicious Over-the-Air (OTA) firmware updates can lead the attacker to take remote control of a fleet of vehicles and adversely affect their expected behaviour [7]. Moreover, a rogue vehicle in V2X can disrupt road traffic or trigger a collision by generating and broadcasting fake safety/traffic data (i.e., wrong traffic prediction information differs from real-world information) over the NGNs

.

Jamming/Blocking attack: Such attacks disrupt wireless networks so that sensors cannot receive messages and local entities (e.g. vehicles) cannot send or receive V2X messages. By blinding or blocking the sensors of the local entities in V2X infrastructure, an attacker can - for example - manipulate an AI model, feed an algorithm with faulty data or intentionally provide scarce data and thus decrease the efficiency of automated decision-making.

B. *Privacy*

5G&B networks will take us much further towards V2X communications, which rely on the sharing of large volumes of often-personal data. It raises an information disclosure issue given the abundance of personal and sensitive information stored and used by vehicles or other local entities for decision-making purposes, including critical data on vehicles. With V2X comes the responsibility of managing the computing infrastructure, which processes huge amounts of data and protecting privacy when collecting sensitive data at the edge. Currently, there is no way to unambiguously determine when linked and de-identified local datasets at nodes (e.g. autonomous vehicles, road side unit, pedestrians etc.) cross the threshold to become personally identifiable [8]. As shown in Fig. 2, there are unresolved issues in V2X communications such as

- How to measure the levels of personal protection required or provided?
- How to define the thresholds for personally identifiable information?
- What factors can be used to quantify personally identifiable information and the level of privacy protection required?
- What data are users/local entities willing to share and with whom, in an understandable way?
- When to measure the level of personal information in a dataset?

- When is an individual reasonably identifiable given rich contextual environments, personal attributes, spatial and temporal aspects of data?

These issues suggest that an agreement among stockholders, policy makers and governing bodies, local authorities, and users is required to tradeoff between managing privacy with building the required trust. V2X communications require the new trust models with updated privacy protection approaches.

IV. Case Study

We present a case study in which we have designed and conducted an experimental analysis of a data exchange model for autonomous vehicles in V2X communications. V2X require incredible data processing capabilities and speeds needed to mimic the timing of human reflexes. Edge computing will enable lightning-fast response time because of 5G&B's promise of lower latency, ability to offload computing tasks, and better location awareness. However, when it comes to edge computing, there are many challenges in terms of network trustworthiness [9]. For example, an automotive control system consists of an in-vehicle network that connects all the devices, monitors the state of the automatic transmission of the vehicle engine, and manages the sensors inside the vehicle. While exchanging real-time traffic updates and road hazard information, which makes driving safer and more efficient, an attacker can launch attacks such as man-in-the- middle attack, bogus information attack, DoS, location tracking, malicious codes, and replay attacks [10]. With autonomous vehicles comes the responsibility at network architecture level for managing the computing infrastructure, which processes massive amounts of data and privacy protection when collecting sensitive data at the edge [11].

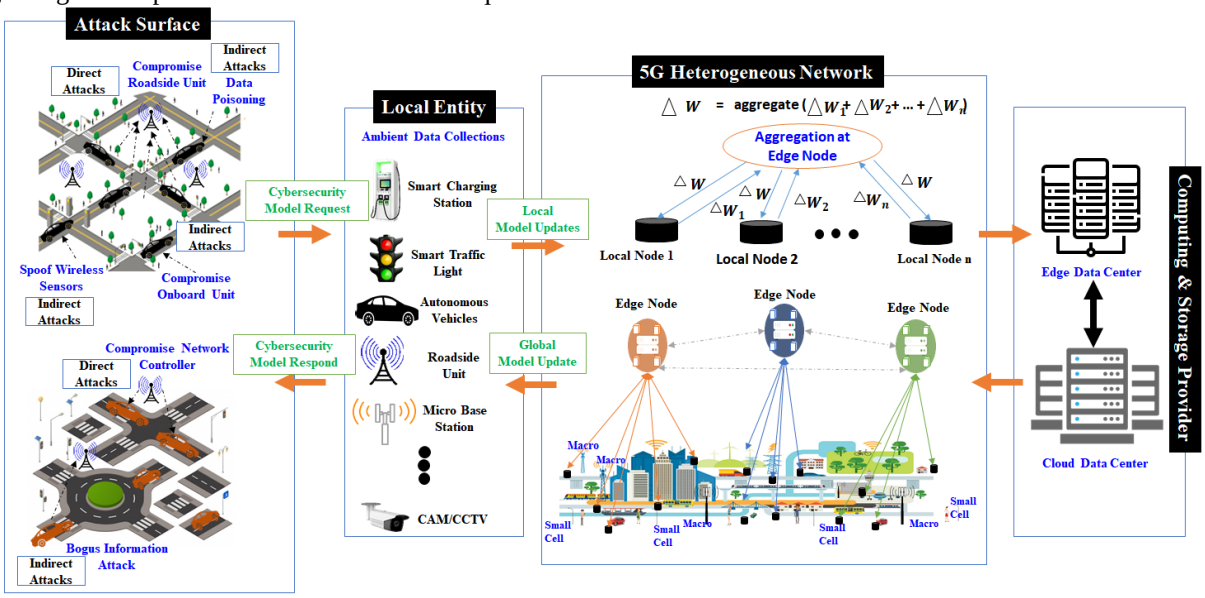
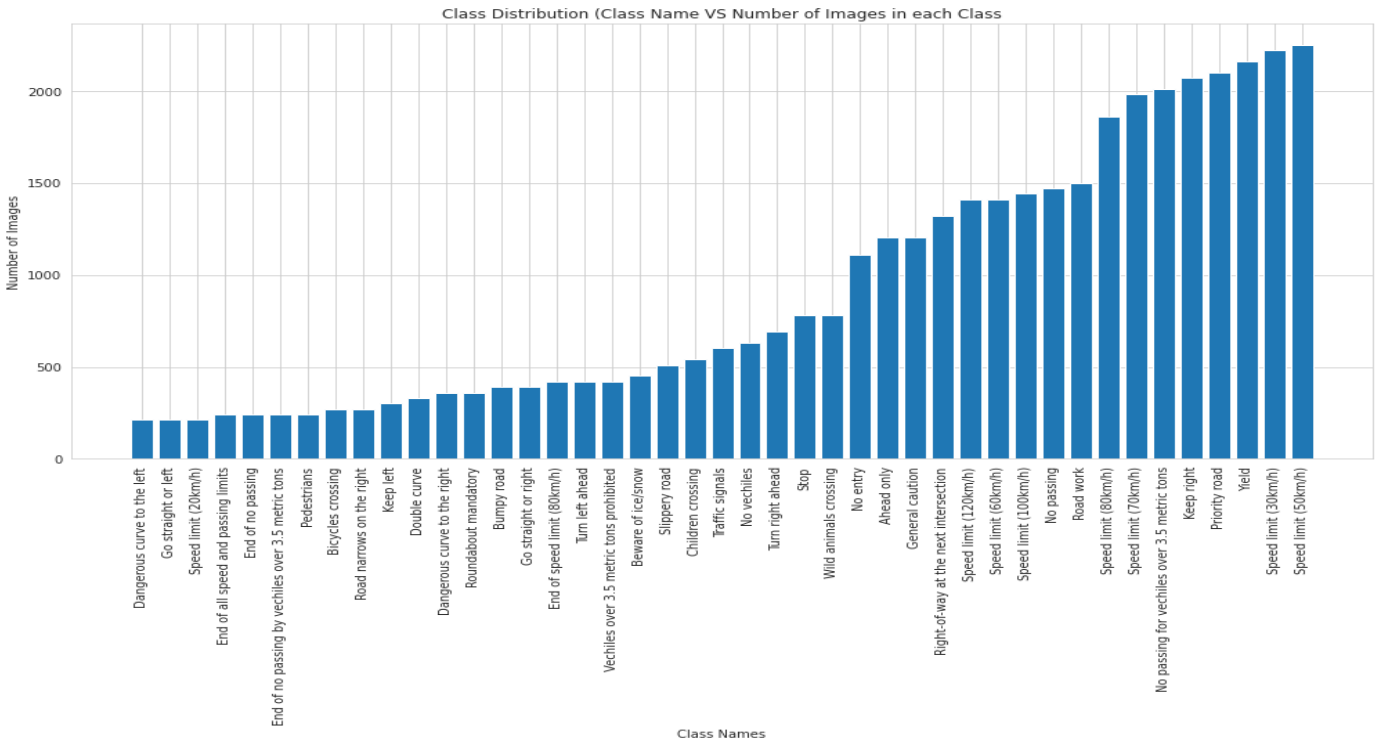


Fig. 3 Abstract overview of the data exchange model

A. Models

As depicted in Fig. 3, to assess the feasibility and effectiveness, we have designed prototype for a data exchange model using collaborative learning that allows the system to perform processing on the device without sharing its dataset with an intermediate nodes/server. Such learning pushes data processing to the network's edge and enables local nodes to collaboratively learn a shared model using local data on node and keeping the data local. In turn, this approach significantly reduces privacy leakage risks. Federated learning [12] and Split learning [13] are two collaborative learning models that we have used which allow for training a model collectively from various distributed data sources without sharing raw data. Federated learning is more efficient by increasing the number of data samples, especially when the number of clients or model size is small, while split learning is efficient with an increasing number of clients and highly scalable with a number of model parameters [14]. Data at local nodes (e.g. vehicles) contains valuable information and these data become the key to building a trusted model for autonomous vehicles to provide personalized services to maximally improve the user experience in a secure manner via V2X communications.

To assess the feasibility, in this case study, we split the convolutional neural network (CNN) model into parts i.e. local node and aggregator edge node parts. Similar to federated learning, we train the global model in the split learning without sharing the actual data. We also tried with segments the networks in split learning and each being trained on a local node with a segment. Moreover, all the segments at the local nodes communicate with each other with their hidden states to train the model as well as preserve the confidentiality of the temporal relationship between the segments at each local node. However, this approach has a limitation due to the high mobility of local nodes in V2X communications. Keeping the potential requirements of intelligent services and applications in next generation networks, the high mobility factor is a crucial aspect in realizing all V2X communications solutions.



(a) Class distribution

(b) Image variation

Fig. 4 Distribution of the images and its variation

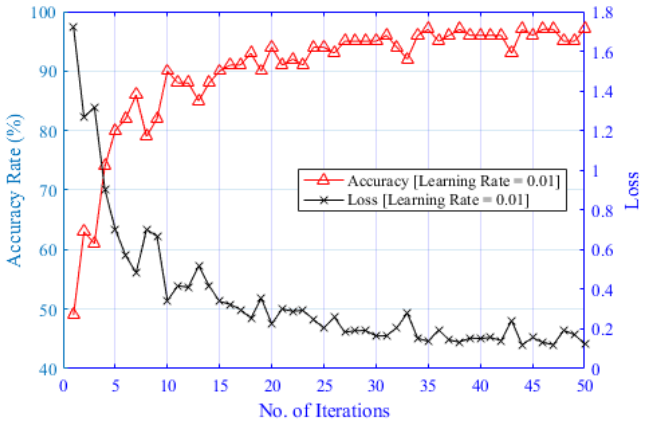
B. *Datasets*

In this case study, we have used the German Traffic Sign Recognition Benchmark (GTSRB) dataset [15] to train the models using federated and split learning in a real world scenario. The dataset is significantly unbalanced, reflecting the real world scenario where each client has significant variation in the number of images to classify for various classes. The dataset consists of a single-image, multi-class classification problem with over 40 classes and 50,000 images in total. In addition, the images are distributed unevenly among these classes. Fig. 4 shows that the distribution of the images is not uniform and that the dataset has problems of class imbalance. There is a huge disparity in the data, e.g., the most well-represented class (Speed limit 50km/h) contains nearly 2,000 cases, while the least well-represented class (Speed limit 20km/h) has only 200. In addition, the images are substantially different in terms of contrast and brightness. Humans are unable to comprehend and categorize some of these signs since they are in complete darkness. This will be the case with autonomous vehicles in a real world scenario where each local node will have its own data and class distribution.

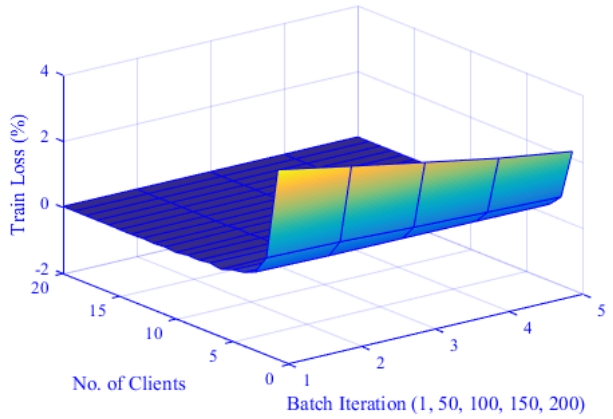
C. *Experimental analysis*

We have implemented both federated learning and Split learning models on the GTSRB dataset to assess the feasibility of the data exchange model in autonomous vehicles. We have installed Pysyft version 2.9 in Colab notebooks, a Python library that decouples local dataset from model training. In the federated learning, we created 20 clients and randomly distributed the dataset among those clients. Here, we have assumed that all clients will have all the classes but the distribution of the number of images will be non-uniform. We have used PyTorch to create virtual workers where each workers is considered a client (i.e. local node).

Fig. 5 (a) and (b) show the result of accuracy rate achieved by both federated leaning and split learning techniques. In Fig. 5 (a), the results indicate that the aggregated model achieved 0.961 accuracy rate in 50 iterations. While the aggregated model in Split learning achieved 0.976 accuracy rate in 20 iterations, as shown in Fig. 5 (b). We repeated the steps 20 times to check the accuracy of the result and observed the average accuracy achieved. Table 1 presents the values of the parameters in the experimental analysis. We also observed the accuracy of the GTSTRB dataset using a CNN model without distributing the dataset and splitting learning model, and achieved an accuracy rate of 0.984. Based on the our observation in the case study, we have noticed that the collaborative learning technique is a feasible solution for privacy-preserving data exchange in V2X communications without degrading the accuracy, and sharing or aggregating (at central node) the data from local nodes.



(a) Federated learning model



(b) Split learning model

Fig. 5 Result of accuracy and loss achieved by both federated leaning and split learning models

D. Discussion

As shown in Table 1, we observed significant changes in accuracy rates with the variations in epochs, learning rates for different models. With a smaller number of epochs and an appropriate learning rate, the CNN model achieved highest accuracy considering single client and central data store. In contrast, with the same number of clients, batch size, and learning rate, the split learning model obtained slightly higher rate of accuracy than federated learning model. Thus, split learning model outperforms over both models of CNN and federated learning. To best of our knowledge, the split learning approach is limited to CNN and recurrent neural networks (RNN). When we talk about intelligent services in next-generation networks, we need to extend existing solutions to more "unconventional" NNs, such as Transformer Capsule Networks. Additionally, we can provide solutions with built-in privacy-preserving technologies such as secure multi-party computations, zero-knowledge proofs while sharing collaborative learning model parameters in V2X communications. This research direction will provide new results and sustainable solutions with focus on security and privacy in V2X communications.

Table 1: Parameters

Model	No. of Clients	Batch Size	Epochs	Learning Rate	Accuracy
CNN	-	128	10	0.01	0.969
				0.02	0.984
Federated Learning	20		50	0.01	0.953
				0.02	0.961
Split Learning			200	0.01	0.958
				0.02	0.976

V. FUTURE IMPACT

V2X communications for the automotive industry are a vital economic driver in the world. As the entire automotive industry is moving towards e-mobility and self-driving cars, this work will play a key role in determining how far we are from the reality of smart mobility and increase competitiveness of the sector. In the coming years, the complete evolution of the tech-driven automotive industry is expected to revolve around security, safety of the driver, passengers, and vehicles, and this work provides an idea to build the foundation for stakeholders working on the future of V2X communications features of the new-age vehicles. Cyber resilience in V2X communications is of utmost importance. The system should have the ability to prepare for, react to and recover from cyber-attacks. However, security and privacy issues complicate the implementations of autonomous driving. By using collaborative learning, as a "user" becomes more dependent on the circumstances of digital driving, the system becomes more familiar with that user's locations, behaviors and habits. Exploiting this information is critical to provide reliable autonomous driving services in V2X communications. At the same time, unless the use of such information is strictly controlled, disclosure of a user's information to unknown sources can lead to identity theft. This study identifies insightful aspects for future V2X communications that are worth considering are:

Network node dangers: Connected and autonomous vehicles are the key nodes within all over the network of V2X. These key nodes use collaborative learning to smart decision, including identifying traffic signs, predicting road traffics efficiently. However, an attacker can hack a node, which can lead to falsify whole collaborative learning model. Further, the false model can mislead every node or vehicle in the network. The traditional wireless node security mechanism, including malicious node detection, node attestation is not appropriate to deploy due to unique requirements of V2X network such as ultra-low latency and ultra-connectivity. Therefore, an appropriate node security to fully deployment of collaborative learning and prevent manipulation or falsification.

Ransomware threat: Ransomware can be a potential threat to autonomous driving service in V2X communications. It can significantly risk to a data exchange model (i.e., discussed in A) wherein cyber-criminals can hostage a service provider or a driver and lock data and systems with the intent of blackmailing them for a ransom. The risk can significantly halt exchange of learning data within the entire network which further decrease in traffic safety. Thus, development of potential solutions and awareness to prevent from a ransomware attack are the key

envisioned by which trusted exchange of learning data could be made.

Secure transmission: All the data transmission in V2X should be encrypted in order to protect privacy and the integrity of learning data being sent back and forth between the vehicles. Moreover, all the endpoints and infrastructure, including drivers and road safety equipment's should be transparent and employ defense mechanism that are resistant to potential security threats. The deployment of blockchain technology could be envisioned to facilitate a transparent and encrypted data exchange between the entities in the V2X communication. It also supports the recording the data transaction and tracking assets in the V2X network in the form a shared, distributed and immutable ledger.

Security standardization: The uniform security standards need to be adopted when implementing autonomous driving services in V2X communications to tighten the security of connected and autonomous vehicles. A Self-regulation is not sufficient and a huge paradigm shift in the security of collaborative learning is required in the future as everything is connected and advancement in communication technology with 5G, and one day 6G.

VI. Conclusion

Today's automakers face certain challenges, such as how to trust the information received by each vehicle; how to achieve consistent and reliable communications between vehicles, local base station, infrastructure and network under complex and arbitrary environmental conditions. V2X communication is a critical factor in the success of the automotive industry. This study provided an overview of critical requirements of the network architecture, discussed the open security and privacy issues in V2X communications. Based on the case study, we observed that using collaborative learning techniques could be potential solution to build secure and privacy-preserving data exchange model in V2X communications. In the future, we will explore collaborative learning to quantify privacy issues in V2X communications in more detail. In addition, we will also explore how cutting-edge technologies such as Blockchain, Software-defined Networking (SDN) can be integrated with collaborative learning to provide a robust, trustable and cyber-resilient platform in V2X communications.

Acknowledgement

We thank Professor Nir Oren (Department of Computing Science, University of Aberdeen, UK) and Professor Steven Furnell (University of Nottingham, UK) for their expertise and assistance throughout all aspects of our study and for his contribution to the technical review and proofreading the article.

References

[1] Posner, J., Tseng, L., Aloqaily, M., & Jararweh, Y. (2021). Federated learning in vehicular networks: opportunities and solutions. IEEE Network, 35(2), 152-159

[2] Posner, J., Tseng, L., Aloqaily, M., & Guizani, M. (2020, November). Federated Vehicular Networks: Design, Applications, Routing, and Evaluation. In 2020 IEEE 45th Conference on Local Computer Networks (LCN) (pp. 429-432). IEEE

[3] Bourdoux, A., Barreto, A. N., van Liempd, B., de Lima, C., Dardari, D., Belot, D., ... & Suutala, J. (2020). 6G White Paper on Localization and Sensing. *arXiv preprint arXiv:2006.01779*

[4] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. [Avaiable Online]: https://www.sae.org/standards/content/j3016_201806/ (Accessed on 12th Sep 2021)

[5] Alnasser, A., Sun, H., & Jiang, J. (2019). Cyber security challenges and solutions for V2X communications: A survey. *Computer Networks*, *151*, 52-67

[6] Ghosal, A., & Conti, M. (2020). Security issues and challenges in V2X: A Survey. *Computer Networks*, *169*, 107093

[7] Bauwens, J., Ruckebusch, P., Giannoulis, S., Moerman, I., & De Poorter, E. (2020). Over-the-air software updates in the Internet of Things: An overview of key principles. IEEE Communications Magazine, 58(2), 35-41

[8] Ylianttila, M., Kantola, R., Gurtov, A., Mucchi, L., Oppermann, I., Yan, Z., ... & Röning, J. (2020). 6g white paper: Research challenges for trust, security and privacy. arXiv preprint arXiv:2004.11665

[9] Liu, S., et al. (2019). Edge computing for autonomous driving: Opportunities and challenges. Proceedings of the IEEE, 107(8), 1697-1716

[10] Kim, K., Kim, J. S., Jeong, S., Park, J. H., & Kim, H. K. (2021). Cybersecurity for Autonomous Vehicles: Review of Attacks and Defense. *Computers & Security*, 102150

[11] Lai, C., Lu, R., Zheng, D., & Shen, X. S. (2020). Security and privacy challenges in 5G-enabled vehicular networks. IEEE Network, 34(2), 37-45

[12] McMahan, H. B., Moore, E., Ramage, D., and y Arcas, B. A. (2016). Federated learning of deep networks using model averaging. CoRR,http://arxiv.org/abs/1602.05629

[13] Vepakomma, P., Gupta, O., Swedish, T., & Raskar, R. (2018). Split learning for health: Distributed deep learning without sharing raw patient data. arXiv preprint arXiv:1812.00564

[14] Singh, A., et al. (2019). Detailed comparison of communication efficiency of split learning and federated learning. arXiv preprint arXiv:1909.09145

[15] German Traffic Sign Recognition Benchmark. [Available Online]: https://benchmark.ini.rub.de/?section=gtsrb&subsection=dataset (Accessed on 12th Sep 2021)

Pradip Kumar Sharma [M'18, SM'21] (pradip.sharma@abdn.ac.uk) is an assistant professor of cybersecurity in the Department of Computing Science at the University of Aberdeen, United Kingdom. His research interests are in the areas of cybersecurity, Blockchain, edge computing, SDN, Privacy-aware AI, and IoT security.

Deepanshu Vohra is currently pursuing an MSc AI at the University of Aberdeen, UK. Prior to pursuing his master's degree, he spent more than ten years in the Asset Management and Wealth Management divisions of some of the world's largest financial institutions such as JP Morgan, Goldman Sachs, Fidelity Investment.

Shailendra received a Ph.D. degree at Seoul National University of Science and Technology, Seoul, South Korea. His broad research interests include Cyber Security, IoT Security, Applied Artificial Intelligence, and Blockchain. He has published his research outcomes in various top tier international Q1 journals, including IEEE transactions and magazines.