Heudiasyc
Vulnerable Agents Safety in Cloud Hosted Intelligent Transportation Systems: Literature Review

João Pedro Sandrini Milanezi

Alessandro Correa Victorino, Tereza Cristina Melo de Brito Carvalho and Moisés Renato Nunes Ribeiro

November 26, 2025

Acknowledgments

The opinions, hypotheses, conclusions, and recommendations expressed in this material are the sole responsibility of the author and do not necessarily reflect the views of FAPESP. This study was financed, in part, by the São Paulo Research Foundation (FAPESP), Brazil. Process Number 2025/14162-0.
I would also like to thank, once again, my dear Professors Tereza Cristina Melo de Brito Carvalho, Moisés Renato Nunes Ribeiro and Alessandro Correa Victorino who made this exchange program possible.

Contents

1 Introduction **3**

2 Vulnerable Agents Safety **3**
2.1 Road Vulnerable Agents 3
2.2 AV Vulnerable Agents 3

3 Cloud-Edge for Critical Applications **4**
3.1 Edge Processing in ITS 4
3.2 Performance Isolation 5

4 Proposed Work **6**

5 Conclusion **7**

1. Introduction

In the scope of autonomous Intelligent Transportation Systems (ITS), and in the context of autonomous vehicles (AV), semaphore systems, traffic intersection models or any other, safety regarding vulnerable agents, like drivers, passengers and pedestrians is a real concern that must be addressed in ITS [1].
This work tries to aggregate relevant research regarding vulnerable agents in ITS, cloud hosted ITS and performance isolation for cloud and how both subjects can be integrated to improve vulnerable agent safety. We also propose a research project based on the reviewed literature and discussed topics.

2. Vulnerable Agents Safety

2.1. Road Vulnerable Agents

V. M. G. Martinez [2] proposes an ontology-based Digital Twin (DT) for road intersections, and explicitly addresses vulnerable road users (VRU) safety.
The author points as the main challenge in the current development of Digital Twins the understanding data and flows to represent more accurately the physical space in both it's static and dynamic aspects.
Ontology has a fundamental role in developing accurate DT and are also contributes to VRU safety. With that, a systematic framework for the development of an ontology for road intersection DT. The framework is comprised of four layers: Physical, Data, Model and Application.
Following, the authors showed a complex information model developed in a way that can be implemented using technologies such as Next Generation Service Interfaces Linked Data (NGSI-LD) and the open-source smart solutions framework, FIWARE. The information model treats dynamic and static data separately and also model traffic, connecting static and dynamic data to predict scenarios and prevent accidents.
Lastly, the authors demonstrate how the proposed methodology can be used to create a DT, instantiating a completely open-source DT based on it, implementing the information model using the FIWARE Framework and simulating a traffic intersection scenario.

2.2. AV Vulnerable Agents

Regarding safety, the passengers of VA are also a concern, as for example, motion comfort needs to be addressed in the case of passengers with specific needs such as pregnant women and elderly people. M. Aledhari [3] outlines motion comfort for AV, and how it affects the physical and psychological state of passengers.
The authors expose various methods developed to improve comfort and decrease motion sickness from AV. Among other examples, including a motion planning algorithm, multimodal systems mounted to the AV, a simulation framework for accessing ride comfort shows itself as a great solution [4].
The simulation utilizes road surface model and Monte Carlo simulator to estimate optimal comfort parameters. Utilizing the ISO-2631 to obtain accurate motion sickness and comfort readings. The simulation's parameters are loaded into a XML file, providing flexibility

for different scenarios. MATLAB/Simulink and CarMaker are the simulation softwares used. Finally, the simulations exposed a relation between acceleration and comfort.
Moreover, various applications of methodologies and techniques for comfort optimization are reviewed, including speed control in rough pavements, cruise control systems, path trackers, collision avoidance and predictive control models. All those solutions rely heavily on AI and machine learning to achieve significant ride comfort improvements.
Concerning the adaptability provided by loading the parameters into a XML in the simulation framework and, a significant improvement in both flexibility and result accuracy could be achieved utilizing ontology-based models and smart data models with NGSI-LD as in [2].
Lastly, a smart data model would also provides great integration with various software solutions, as the FIWARE [1] framework enables data interoperability and the possibility to expand the simulation framework into a DT system, enabling better physical representation and prediction [1].

3. Cloud-Edge for Critical Applications

Nowadays, digital infrastructures are being reshaped based in applications each time more demanding and with stricter requirements. While centralized Cloud infrastructures provides cost-effective and bulk resource provisioning, it often fails in reaching critical requirements of latency and decentralization [5].
On the other hand, Edge computing offers the possibility to processes information in a distributed way, closer to the final application. Thus, enabling applications with stricter requirements, such as real-time capable services [5].
With the exposed, Cloud-Edge technologies have gained increasing notability in the recent days. Software like StarlingX [2] with a container based archtechture, native low-latency capabilities, low footprint, TSN and performance isolation techniques native support leverage the Edge-Cloud paradigm, providing powerful resources to address applications critical requirements [6], while being able to operate in resource-constrained environments, precisely, the Edge.

3.1. Edge Processing in ITS

V. M. G. Martinez et Al. [7] proposes a Digital Twin for traffic intersections, as a way to enable real-time control, monitoring and management of intelligent intersections, as well as a tool to improve safety at those intersections. The author's implementation uses an Edge-Cloud infrastructure based on StarlingX as enabler technology to make data and models always accessible and also. Its also mentioned that the low latency and deterministic nature of StarlingX enables the exploration of real-time applications.
Nevertheless, in order to predict scenarios and prevent accidents using the DT platform, High Performance Computing resources are needed, as the decision-making algorithms have high computational requirements. As those systems also needs to achieve low-latency,

[1] https://www.fiware.org
[2] https://www.starlingx.io/

the authors suggest the allocation of resources on the edge instead of regional clouds, to improve communication latency while still having significant computational resources.
L. Bai [8] introduces the concept of collaborative edge intelligence as a way of overcoming both the bottleneck of conventional Clouds, which suffer from unstable and slow communication due to large distance links, and the edge, with limited resources from individual devices. The collaborative edge intelligence use multiple edge nodes to leverage a federated resource pool in a distributed edge and enhance communication with centralized clouds. The authors points as advantages of this paradigm the enlargement of resource pools, low-latency data processing capabilities and bigger area covered by services.
S. Liang [9] proposes an object detection Edge-Cloud system for Autonomous Vehicles. The system uses a lightweight object detection framework that can be deployed on resource-constrained edge devices and still be able to perform well, even when compared to a cloud deployment, obtaining up to 47.3% accuracy (mAP) in the COCO2017 dataset, at 26.6 frames per second, while all parameters take only 25.67 MB.

3.2. Performance Isolation

In cloud infrastructures, the provisioning of resources for multiple applications can result in noisy neighbors affecting other applications and performance losses. Taking this into account, novel isolation techniques such as Kata Containers [3] and Unikernels are emerging[4] to address the isolation problem.
H. Dinh Tuan [10] compares conventional Linux containers with Unikernels in environments where resources are constrained, for edge workloads. Unikernels work by also virtualizing the kernel, which means each container has its own specialized kernel, adding another isolation layer for containers.
The results show that in extreme-edge scenarios, the minimal overhead of Unikernels is crucial to enable maximum application performance, while in systems where memory is the main bottleneck, the stability of conventional containers is essential. Lastly, in resource-rich environments, the deployment choice depends on the applications needs. Thus, the choice between deployments has proved to be an important engineering decision based on system requirements and trade-offs.
Concerning MicroVMs such as Kata Containers and Firecracker [5], K. Lee [11] makes a performance comparison and analysis of applications deployed natively, with docker containers, VMS and MicroVMs, on edge hardware.
The results showed overall advantages in all evaluated scenarios: object detection, file and network I/O and Energy efficiency, while the most significant result was with network I/O, with MicroVMs showing up to 2× better performance and 3.9× for random write operations, when compared to containers. All this corroborates the adoption of MicroVMs in production enviroments, according to the authors.
E. A. S. Junior [5] presents a case study of StarlingX as Containers as a Service (CaaS) native edge infrastructure and compares it with a traditional Infrastructure as a Service (IaaS) OpenStack environment, aiming to understand resource footprint and performance of modern edge platforms.

[3]https://katacontainers.io/
[4]http://unikernel.org/
[5]https://firecracker-microvm.github.io/

The choice of STX is based off of its premise of delivering a low-latency and small footprint edge platform. The authors found that STX enhanced OpenStack (CaaS) performance in Inter-VM output, when compared to a conventional IaaS deployment. Also, intense I/O demands of an application can affect neighboring containers, raising concerns with isolation. Finally, STX showed to be more flexible and to provide resource allocation advantages, although consuming more CPU in idle then IaaS OpenStack.
The author report that choosing a method of deployment involves tradeoffs, but for time sensitive applications on the edge, such as ITS and AV systems, Unikernels present itself as a better choice, providing better isolation with little overhead.
Its also important to note that StarlingX has native support to the open source solution Kata Containers, that work similarly to Unikernels, virtualizing a specialized kernel for each service and provision resources in the form of lightweight virtual machines. That can be further explored to simplify the implementation of isolation techniques in the proposed project.
StarlingX also allows the provisionment of isolated CPU Cores for specific applications, which can be explored as a way of improving application performance in CPU-intensive scenarios. This function allows the system administrator to address CPU Cores exclusively for an application, which means that in scenarios where the server resources are being heavily used, its possible to assure that specific applications always have the resources necessary to achieve optimal QoS. This is particularly useful in the context of critical applications, as it can assures that the critical application will always achieve its requisites, even when the hardware resources are being heavily demanded by other less-important applications hosted on the same infrastructure.

4. Proposed Work

Based on the reviewed literature, available resources at Heudiasyc (UTC) and the BEPE project we propose an experiment to measure how application isolation impacts performance Vehicle-to-Cloud (V2C) communication in edge-cloud environments.The proposed experiment consists of an AV running inside the Heudiasyc internal test circuit and communicating with the private SarlingX edge server deployed locally. Figure 1 clarifies the experiment setup.
On the server side, two instances of an ADAS application-stack should be running, one instance in conventional Linux Containers and one utilizing Kata Containers, and receiving the data from the AV. With V2C communication working, we can measure communication and application delays, thus, measuring how isolation impacts performance. Its also possible to measure resource usage, through the STX metrics server, to access isolation from a lower-level perspective.
This scenarios recreates a real use case where heavier computations loads of AV are transferred to the Edge, that has more computational resources available than in the embedded system, while still being able to achieve low-latencies. More resources available means that its possible to run more complex and accurate AI and Machine Learning models for better navigation, comfort and accident prevention.

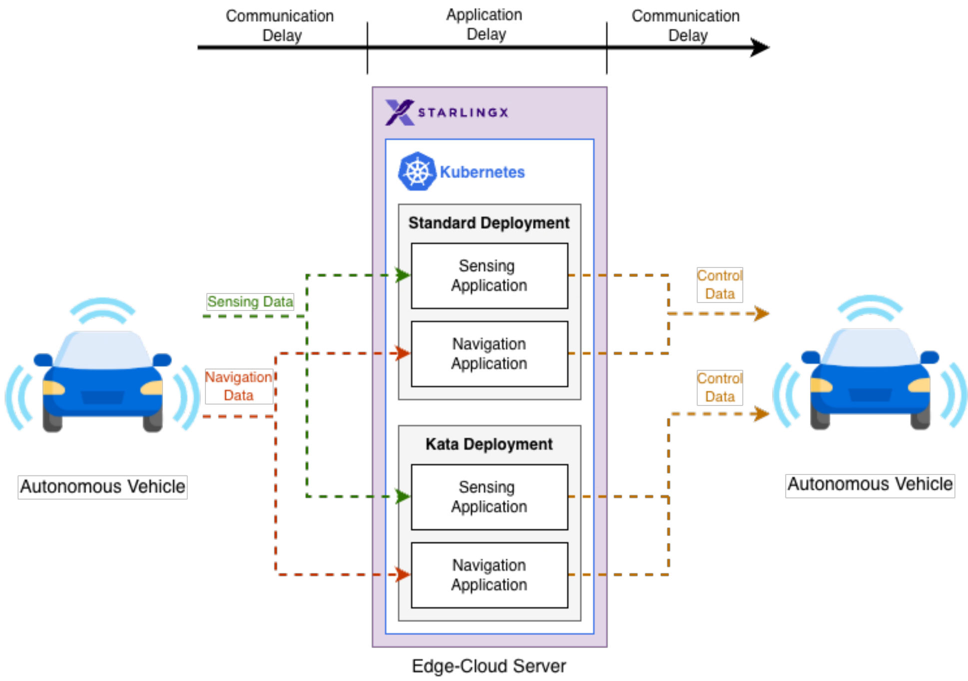

Figure 1: Your figure caption here

5. Conclusion

Vulnerable agents safety and comfort in the context of ITS and AV it's an important subject that still needs improvement in modern systems. Edge-cloud technologies can help improve those aspects by tackling challenges in ITS, like supporting applications to represent more precisely the physical environment in the digital world, while guaranteeing low-latencies and great QoS.

The proposed work represents a step foward on this direction of solving modern ITS problems. Utilizing the Edge-Cloud platform we will be able to demonstrate a real use case of this infrastructure to achieve strict requirements from the application, leveraging the development of more complex, safe and comfortable ITS. Nevertheless, we also aim to demonstrate how isolation can be used in the context of cloud-computing to enhance application performance in various applications, contributing not only to better QoS of those applications, but to optimize energy consumption through efficient usage of computational resources.

References

[1] V. M. G. Martinez, D. R. Campelo, and M. R. N. Ribeiro, "Sustainable intelligent transportation systems via digital twins: A contextualized survey," IEEE Open Journal of Intelligent Transportation Systems, vol. 6, pp. 363–392, 2025.

[2] V. M. G. Martinez, S. S. Mussi, M. R. N. Ribeiro, D. R. Campelo, and T. C. Carvalho, "Vru safety at road intersections: An ontology-driven digital twin," IEEE Access, vol. 13, pp. 138166–138184, 2025.

[3] M. Aledhari, M. Rahouti, J. Qadir, B. Qolomany, M. Guizani, and A. Al-Fuqaha, "Motion comfort optimization for autonomous vehicles: Concepts, methods, and techniques," IEEE Internet of Things Journal, vol. 11, no. 1, pp. 378–402, 2024.

[4] A. Genser, R. Spielhofer, P. Nitsche, and A. Kouvelas, "Ride comfort assessment for automated vehicles utilizing a road surface model and monte carlo simulations," Computer-Aided Civil and Infrastructure Engineering, vol. 37, no. 10, pp. 1316–1334, 2022.

[5] E. A. S. Junior, V. M. G. Martínez, R. C. Mello, M. R. N. Ribeiro, A. W. Donatti, T. C. Carvalho, J. H. Corrêa, R. Moreira, J. M. de Oliveira, J.-M. S. Nogueira, and F. O. Silva, "Edge clouds on open infrastructure: Assessing footprint and performance for iaas over caas," in Proceedings of the Congresso de Inovação e Tecnologia (CIOT 2025), (London, United Kingdom), CIOT, 2025. Accepted and presented, to appear.

[6] Y. Wang, "Time-sensitive networking (tsn) enabling on starlingx." Virtual Open Infrastructure Summit, 2020. [Online]. Available: https://www.youtube.com/watch?v=DgYA5m0d87U.

[7] V. Martínez, M. Ribeiro, and D. Campelo, "Intelligent road intersections: A case for digital twins," in Anais do III Workshop Brasileiro de Cidades Inteligentes, (Porto Alegre, RS, Brasil), pp. 151–158, SBC, 2022.

[8] L. Bai, J. Cao, M. Zhang, and B. Li, "Collaborative edge intelligence for autonomous vehicles: Opportunities and challenges," IEEE Network, vol. 39, no. 2, pp. 52–60, 2025.

[9] S. Liang, H. Wu, L. Zhen, Q. Hua, S. Garg, G. Kaddoum, M. M. Hassan, and K. Yu, "Edge yolo: Real-time intelligent object detection system based on edge-cloud cooperation in autonomous vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 25345–25360, 2022.

[10] H. Dinh-Tuan, "Unikernels vs. containers: A runtime-level performance comparison for resource-constrained edge workloads," 2025.

[11] K. Lee, Y. Choi, and B. Tak, "Performance analysis of microvms and containers for edge computing: A focus on file and network i/o," Future Generation Computer Systems, p. 108176, 2025.