INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 04, Issue 12, Decembaer 2024, pp : 1467-1468
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

DEDUPLICATION IN AUTONOMOUS VEHICLES - ENHANCING EFFICIENCY AND DATA INTEGRITY

Mr. Pradeep Nayak[1], Nischitha[2], Omkar K S[3], Omkarayya Jagadish Kambalimath[4], Pallavi Sadashiv Khadnavar[5]

[1]Faculty, Department of Information Science and Engineering Alva's institute of Engineering and Technology, Mijar, Mangalore, Karnataka, India.

[2,3,4,5]Students, Department of Information Science and Engineering Alva's institute of Engineering and Technology, Mijar, Mangalore, Karnataka, India.

DOI: https://www.doi.org/10.58257/IJPREMS37867

ABSTRACT

Deduplication is a critical data management technique for reducing redundancy, optimizing storage, and ensuring real-time data processing in autonomous vehicles (AVs). This review explores the role of deduplication in AV systems, emphasizing its integration with sensor data, path planning algorithms, and cybersecurity protocols. By analyzing current methodologies and technologies, this paper identifies the challenges and opportunities for implementing deduplication in vehicular systems. The findings highlight how deduplication can significantly enhance efficiency, reliability, and performance in autonomous transportation, paving the way for robust and scalable AV systems.

Keywords: Deduplication, Autonomous Vehicles, Data Integrity, Path Planning, Cybersecurity.

1. INTRODUCTION

Autonomous vehicles (AVs) rely on extensive data from sensors, cameras, and onboard systems to navigate and make decisions in real time. The continuous generation of massive datasets can lead to data redundancy, which not only wastes storage resources but also hinders real-time processing efficiency. Deduplication addresses these challenges by eliminating redundant data while maintaining its integrity. This review paper aims to provide an overview of deduplication techniques, their applications in AV systems, and the benefits they bring to the functionality and reliability of autonomous systems.

2. METHODOLOGY

2.1 Deduplication Techniques in AV Systems

Deduplication techniques can be categorized into file-level, block-level, and byte-level methods. In AV systems, block-level deduplication is particularly effective for handling high-volume data generated by LiDAR, radar, and cameras. Algorithms like Content-Defined Chunking (CDC) and hashing are employed to detect and eliminate redundancy in real-time data streams.

2.2 Integration with AV Architectures

AVs utilize complex architectures involving various sensors and processing units. Deduplication is integrated into these architectures to optimize data storage and transmission. For instance, deduplication algorithms can be applied to compress redundant GPS and path-planning data, ensuring efficient use of onboard memory.

2.3 Algorithms for Route Deduplication

Route deduplication in autonomous vehicles is achieved using clustering and hashing-based algorithms. A commonly employed approach is the Dynamic Time Warping (DTW) algorithm combined with k-means clustering. This method identifies and groups similar routes based on their spatial and temporal characteristics. Another approach involves using hash-based algorithms, where route segments are hashed, and duplicate segments are identified through hash collisions. These algorithms ensure efficient route management, particularly in dynamic urban environments where overlapping routes are common.

2.4 Sensor-Based Deduplication Algorithms

Sensors in autonomous vehicles, such as LiDAR and radar, employ specific algorithms for deduplication of route data. For example, LiDAR sensors utilize voxel grid filtering, which partitions the point cloud into small cubes (voxels) and retains only the centroid of points within each voxel. This reduces redundancy without losing critical spatial information. Radar sensors use the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm to cluster data points and remove overlapping or redundant measurements. These techniques are crucial for ensuring that sensor data is efficiently processed and utilized in real-time decision-making.

3. MODELING AND ANALYSIS

3.1 Sensor Data Deduplication

Sensors like LiDAR, radar, and cameras generate overlapping datasets during vehicle operation. Deduplication techniques are employed to remove redundant data points while preserving critical information. This process improves storage efficiency and accelerates data analysis for navigation and obstacle detection.

3.2 Path Planning Optimization

Path planning relies on dynamic data inputs from various sources. Deduplication ensures that only unique and relevant data points are processed, reducing computational overhead and enabling faster trajectory calculations. Algorithms like SLAM (Simultaneous Localization and Mapping) benefit significantly from deduplication by optimizing map data.

3.3 Cybersecurity and Data Integrity

Deduplication also plays a role in enhancing cybersecurity by reducing the attack surface associated with redundant data storage. Techniques like secure hashing and encryption are integrated with deduplication algorithms to safeguard sensitive data.

4. RESULTS AND DISCUSSION

4.1 Benefits of Deduplication

Deduplication in AV systems leads to significant improvements in storage optimization, data transmission efficiency, and processing speed. By reducing redundancy, AVs can handle real-time data more effectively, enhancing their decision-making capabilities.

4.2 Challenges in Implementation

Despite its benefits, deduplication faces challenges such as computational complexity, latency issues, and the need for robust algorithms that can handle dynamic vehicular environments. Additionally, ensuring the integrity and security of deduplicated data requires advanced encryption and verification techniques.

4.3 Comparative Analysis

Table 1 summarizes the performance of various deduplication techniques in terms of speed, accuracy, and resource efficiency.

Technique	Speed	Accuracy	Resource Efficiency
File-Level	High	Low	Moderate
Block-Level	Moderate	High	High
Byte-Level	Low	Very High	Moderate

5. CONCLUSION

Deduplication is an indispensable tool for optimizing data management in autonomous vehicles. By eliminating redundant data, it enhances storage efficiency, accelerates processing, and improves overall system performance. Future research should focus on developing lightweight and scalable deduplication algorithms tailored for AV environments, addressing challenges like latency and security.

6. REFERENCES

[1] Parekh, D., et al. "A Review on Autonomous Vehicles: Progress, Methods, and Challenges," Electronics, vol. 11, no. 2162, 2022.

[2] Kumar, N., et al. "Bayesian Coalition Game as-a-Service for Content Distribution in Internet of Vehicles," IEEE Internet of Things Journal, vol. 1, pp. 544-555, 2014.

[3] Sommer, F., et al. "Survey and Classification of Automotive Security Attacks," Information, vol. 10, no. 148, 2019.

[4] Dai, Y., Lee, S.-G. "Perception, Planning, and Control for Self-Driving System Based on On-board Sensors," Advances in Mechanical Engineering, vol. 12, 2020