COMP1806 Coursework

1 st Author's Bryan Hernandez Upegui and D:001305722, 2 nd Author's Ben Peverall and ID: 001314121, 3 rd Author's Mahammad Isgandarzada and ID:001352304, 4th Author's Neebithan Baskaramoorthy and ID: 001297425 and 5th Author's Luis Alexander Yunga Mendoza and ID: 001141347

I. Introduction

"The Autonomous vehicles (AVs) represent the future of transportation, using Machine Learning (ML) models to navigate, avoid obstacles, and make decisions in real time (Kukkala et al., 2022).[1] "We picked this use case because we all share a common interest in autonomous vehicles as they are being developed as we are learning to drive. This caused us to consider their safety and reliability to make an educated decision whether they are secure enough to use confidently without worrying about potential threats. "Autonomous vehicles represent the future of transportation, utilizing AI and ML to make real-time decisions in complex environments" [2](Autonomous Vehicles: Evolution of AI and Learning Algorithms, 2024). As these vehicles become more prevalent, the need to secure their ML systems from potential attacks has grown increasingly urgent. The vulnerabilities of ML-based systems make them a target for adversaries who can compromise safety by attacking these models. To understand how to protect these systems, it is necessary to explore various attack strategies and their corresponding defense mechanisms. In this report, we investigate the security of ML in autonomous vehicles, focusing on potential attack methods and the ways to defend against them. By understanding these attack vectors, we can develop strategies to safeguard the operation of AVs, ensuring the safety of passengers and other road users.

II. Use case Scenario and Background

The use case scenario involves a fleet of (AVs) autonomous vehicles equipped with sensors (like for example LIDAR, cameras, and GPS) and ML models to detect obstacles, plan routes, and make real-time driving decisions. "The ML models in these vehicles are trained on massive datasets to identify[3](NHTSA, n.d.)." Data from sensors is continuously fed into the system, allowing the vehicle to make decisions autonomously. However, AVs face significant challenges when it comes to security. Their dependence on ML models makes them vulnerable to attacks aimed at manipulating the model's input or corrupting the training data. These vulnerabilities could result in catastrophic outcomes, such as accidents or loss of vehicle control. In this context, understanding attack vectors like **model poisoning**, **adversarial attacks**, and **sensor spoofing** is crucial.

III. Analysis and Discussion

---

[1] **Kukkala, V.K., Thiruloga, S.V. and Pasricha, S.** (2022) Roadmap for Cybersecurity in Autonomous Vehicles. *IEEE Consumer Electronics Magazine*, pp. 1–1. doi: 10.1109/mce.2022.3154346.

[2] Garikapati, D. and Shetiya, S.S. (2024) 'Autonomous Vehicles: Evolution of Artificial Intelligence and Learning Algorithms', *arXiv*. Available at: https://arxiv.org/pdf/2402.17690 (Accessed: 12 November 2024).

[3] **NHTSA** (n.d.) Automated vehicles for safety. *National Highway Traffic Safety Administration*. Available at: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety (Accessed: 10 November 2024).

Attack Team:

Model Poisoning Attacks: As attackers we can use poison attacks to manipulate the data used to train the ML models by injecting incorrect or malicious data, leading to erroneous outputs during the vehicle's operation. There are numerous well-known ways to do this to vehicles for instance, "poisoning the vehicles camera ML image dataset with mislabeled images can cause the vehicle to misidentify objects such as pedestrians or traffic signs [4](Chi et al., 2024) ". One article backs this up by saying "A different class of attacks has gained popularity since 2018, mainly targeting the ADAS systems and the onboard sensors used for perception. Eykholt generated various robust visual adversarial perturbations to a stop sign that resulted in it being misidentified as a 45-mph speed limit sign."[5] This can lead to a multitude of different problems for the cars making them extremely unsafe virtually shelving all self-driving cars using this dataset until it can be resolved.

- **Adversarial Attacks (Enviromental)**: These involve crafting slight modifications to inputs, like adding noise to an image, that deceive the model. In an AV, this could involve an attacker placing small stickers on road signs, causing the ML model to misinterpret a stop sign as a yield sign. Another method for doing this attack is using light sources, such as laser pointers, or altering the environment (e.g., fog generators) to disrupt sensor interpretation. This type of attack is particularly effective because the manufacturer can't fix the problem as easily as technically there is nothing wrong with the cars systems as they are correctly assessing the information it is sensing unaware that it has been tampered with, this can make this type of attack particularly devastating even though it is relatively easy to do.

- **Sensor Spoofing**:
  Spoofing attacks involve feeding false signals to the vehicle's sensors (e.g., GPS spoofing, Radar spoofing, Ultrasonic sensor spoofing). Firstly, GPS spoofing is done when attackers broadcast fake GPS signals near the vehicle to override genuine GPS data. This tricks the vehicle into misinterpreting its true location, potentially causing it to follow incorrect routes, veer off-course, or enter restricted areas. Secondly attackers can do Radar spoofing by sending ingenuine radar signals to manipulate the car's radar sensors into detecting nonexistent objects or missing real obstacles. The vehicle may then abruptly brake for "phantom" obstacles or follow dangerously close to actual vehicles, leading to safety risks. Lastly, Attacks do Ultrasonic sensor spoofing by emitting ultrasonic waves at similar frequencies to interfere with the vehicle's sensors, typically used for close-range detection and parking. This can then affect the car's ability to detect nearby objects can be distorted, leading to errors in parking or collision avoidance at low speeds.

---

[4] **Chi, L., Msahli, M., Zhang, Q., Qiu, H., Zhang, T., Memmi, G., and Qiu, M.** (2024) Adversarial Attacks on Autonomous Driving Systems in the Physical World: a Survey. *IEEE Transactions on Intelligent Vehicles*, pp. 1–22. doi: 10.1109/tiv.2024.3484152.

[5] Kukkala, V.K., Thiruloga, S.V. and Pasricha, S. (2022). Roadmap for Cybersecurity in Autonomous Vehicles. IEEE Consumer Electronics Magazine, pp.1–1. doi:https://doi.org/10.1109/mce.2022.3154346.

- **Backdoors Attacks:** involve adding a hidden flaw to an autonomous vehicle's (AV) model during training. Attackers insert a specific trigger, like an object, color, or pattern, that causes the vehicle to act incorrectly when encountered in the real world. These attacks are particularly dangerous because they remain inactive until triggered, allowing the vehicle to function normally in other situations.
  As attackers could add a backdoor trigger by introducing images of people wearing a particular type of clothing (e.g., bright white hoodies) that causes the AV to ignore pedestrians dressed in this way. This would lead to AV potentially failing to yield or stop for certain individuals, creating serious safety hazards for pedestrians (Wu et al., 2023).[6]

- **Sensor Desynchronization Attacks:** occur when attackers disrupt the timing coordination between the different sensors on an autonomous vehicle (AV). AVs rely on multiple sensors; like LIDAR, radar, camaras, and ultrasonic sensors, working together to perceive the environment and make decisions.
  As attackers could use electronic interference to create delays in response times of different AV sensors, leading to timing mismatches in sensor fusion. This could result in incorrect decision-making, such as underestimating a vehicle´s speed when merging lanes, causing unsafe merges (Li et al.,2021) [7]

Defense Team:

Adversarial Training: To defend against adversarial attacks, AVs can be trained with adversarial examples, which improves the model's robustness in recognizing altered inputs. Attackers may alter GPS signals, LIDAR data, or photos to fool the car into misidentifying things or misinterpreting its environment, but by incorporating all of these examples into the training process, we can fend off the attacks. When paired with frequent retraining, multi-modal sensor fusion can assist in identifying sensor discrepancies, and adversarial training guarantees that AVs can resist changing threats.

- **Data Validation**: By confirming the legitimacy of sensor data, identifying irregularities, and guaranteeing consistency among inputs, data validation protects Avs from attacks. These methods include cross-validation of data by comparing LIDAR, camera, and GPS inputs, cryptographic signature verification, and the identification of anomalous patterns. These protections are strengthened by frequent upgrades and testing, which guarantees that Avs rely on reliable data and lowers their susceptibility to adversarial attacks

[6] Wu, Y., Gu, Y., Chen, Y., Cui, X., Li, Q., Xiang, Y., Tong, E., Li, J., Han, Z., & Liu, J. (2023). Camouflage Backdoor Attack against Pedestrian Detection. Applied Sciences, 13(23), 12752. https://doi.org/10.3390/app132312752

[7] Ao Li, Jinwen Wang, and Ning Zhang. 2021. Chronos: Timing Interference as a New Attack Vector on Autonomous Cyber-physical Systems. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21). Association for Computing Machinery, New York, NY, USA, 2426–2428. https://doi.org/10.1145/3460120.3485350

- **Multi-Sensor Fusion**: A defensive strategy involves using multiple redundant sensors to verify incoming data. For example, if the GPS sensor is spoofed, the vehicle can cross-check the information with data from the LIDAR and camera sensors to detect inconsistencies. The most dependable sensors are given priority in certain situations by dynamic weighting, and complementary sensors make it more difficult for attackers to compromise numerous inputs at once (e.g., LIDAR in poor visibility). To improve resilience to attacks, multi-sensor fusion integrates inputs and ensures temporal and contextual consistency.

- **Creating a unique password:** Autonomous vehicles (AVs) can be protected by using a unique password to secure communication networks, stop unwanted access to sensor data, and guarantee secure firmware updates. It defends against network-based attacks, data manipulation, and remote hacking by limiting access to systems and encrypting important data. By protecting training data and update procedures, unique passwords also guard against model poisoning and data corruption by guaranteeing that only trustworthy parties can alter vital systems. This multi-layered security strategy lowers the possibility of attacks and guarantees the integrity of the vehicle's functionality. For example, in 2019, a hacker gained access to hundreds of cars by merely guessing the GPS tracking applications' default password, which was "123456."

- **Understand your automobile:** While it may be simple to identify issues with a typical car, the tremendous degree of complexity behind autonomous vehicles might make it challenging to identify weaknesses. As a result, before operating an autonomous car, drivers need to become acquainted with their vehicle. By understanding how sensors (such as cameras, LIDAR, and GPS) work, you can spot irregularities like hostile image alteration or GPS spoofing. With this information, sensor integration is optimized, guaranteeing redundancy and consistency across data inputs. Human supervision, frequent software upgrades, and customized security procedures strengthen the car's defenses. The AV can react to threats efficiently by utilizing system-specific insights, guaranteeing dependable and secure functioning.

- **First**, **consider security**: Before integrating apps into cars, automakers should make sure that app developers are fully committed to ensuring strict adherence to industry standards, regular updates, continuous monitoring, rigorous security protocols, and extensive testing. This is because unsafe apps can result in third-party attacks, data breaches, unauthorized access, compromised vehicle systems, privacy violations, and serious risks to the safety of both drivers and passengers.

- Attack Graphs

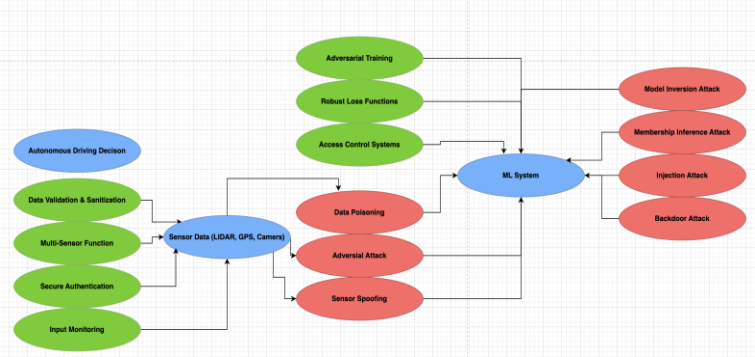

Figure 1

The attack graph (figure 1) shows the possible routes an opponent could take to undermine an autonomous vehicle's (AV) ability to function. A network weakness, like an unprotected Wi-Fi module, allows entry, and it then moves on to internal network exploitation and sensor manipulation. A particular attack vector, like GPS spoofing, is represented by each stage. This can result in compromised sensor outputs, which can then cause the AV to misroute or collide. The graph includes mitigation measures that are in line with each stage of an assault to combat these dangers. For instance, secure Wi-Fi authentication prevents unauthorized network access, while multi-sensor fusion cross-verifies inputs from GPS, LiDAR, and cameras to detect discrepancies. By mapping both attack pathways and corresponding defenses, the graph highlights the interplay between vulnerabilities and mitigations, providing a clear framework for securing AV systems against potential threats.

IV. Conclusions

In this report, we explored various attack vectors targeting ML systems in autonomous vehicles and discussed effective defense strategies. The vulnerability of AVs to adversarial and poisoning attacks demonstrates the need for robust security measures in the design and deployment of these systems. Future work could focus on the development of more sophisticated defense mechanisms, such as real-time anomaly detection, and further research into the integration of machine learning with traditional security practices.

V. Team Contribution

A summary of final credits of member contribution

#	Name	ID	Task given(section)	Cumulative credit
1	Hernandez Upegui, Bryan (Team Leader)	001305722	Introduction, use case Scenario and Background, and attack team	100
2	Neebithan Baskaramoorthy	001297425	Defence Team (work with another teammate to make defending plan), and attack team graph	100
3	Luis Alexander Yunga Mendoza	001141347	Attack Team (work with another teammate to make attack plan)	100

4	Mahammad Isgandarzada	01352304	Defense Team (work with another teammate to make defending plan)	100
5	Ben Peverall	00131412 1	Attack Team, (work with another teammate to make attack plan)	100

References

- **NHTSA** (n.d.) Automated vehicles for safety. *National Highway Traffic Safety Administration*. Available at: https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety (Accessed: 10 November 2024).
- *Citation*: Garikapati, D. and Shetiya, S.S. (2024) 'Autonomous Vehicles: Evolution of Artificial Intelligence and Learning Algorithms', *arXiv*. Available at: https://arxiv.org/pdf/2402.17690 (Accessed: 12 November 2024).
- **OpenAI** (n.d.) use to help make the scenario more in detail and give ideas Available at: ChatGPT (Accessed: 10 November 2024).
- Ornes, S. (2020). *Six Ways to Protect Against Autonomous Vehicle Cyber Attacks*. [online] IEEE Innovation at Work. Available at: https://innovationatwork.ieee.org/six-ways-to-protect-against-autonomous-vehicle-cyber-attacks/.
- **Kukkala, V.K., Thiruloga, S.V. and Pasricha, S.** (2022) Roadmap for Cybersecurity in Autonomous Vehicles. *IEEE Consumer Electronics Magazine*, pp. 1–1. doi: 10.1109/mce.2022.3154346.

Chi, L., Mounira Msahli, Zhang, Q., Qiu, H., Zhang, T., Memmi, G. and Qiu, M. (2024). Adversarial Attacks on Autonomous Driving Systems in the Physical World: a Survey. IEEE Transactions on Intelligent Vehicles, [online] pp.1–22. doi: https://doi.org/10.1109/tiv.2024.3484152

- **Wu, Y., Gu, Y., Chen, Y., Cui, X., Li, Q., Xiang, Y., Tong, E., Li, J., Han, Z., & Liu, J**. (2023). Camouflage Backdoor Attack against Pedestrian Detection. Applied Sciences, 13(23), 12752. https://doi.org/10.3390/app132312752

Ao Li, Jinwen Wang, and Ning Zhang. 2021. Chronos: Timing Interference as a New Attack Vector on Autonomous Cyber-physical Systems. In Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security (CCS '21). Association for Computing Machinery, New York, NY, USA, 2426–2428.

https://doi.org/10.1145/3460120.3485350