Collection, utilization, protection and compliance governance of personal data in the vehicle in the development of auto-drive system

Yineng Xiao [a, b, *]

[a] The Global Intellectual Property Institute, Nanjing University, Suzhou, 215163, China;
[b] Advanced Institute of Information Technology, Peking University, Hangzhou, 311200, China.

ABSTRACT

Vehicle user information and big data are already inseparable, so rapid data processing and analysis has become a major challenge for the current Internet of Vehicles. It is also imperative to study the collection, utilization, protection and compliance governance of personal data in the vehicle during the development of the auto-drive system. This paper analyzes the collection and utilization, protection and compliance governance of personal data in the vehicle during the development of auto-drive system. This paper makes a comparative analysis of the speed of personal data collection and the security of data. Through Hive data collection survey, this paper analyzes the personal data of 2000 different types of car owners in City A, and constructs an experiment from the speed of personal data collection in the car and the degree of personal data protection. The results show that Hive's data collection speed is far faster than that of traditional methods in any aspect. The use of new data protection system can not only reduce the number of external attacks by about 33% through its powerful functionality, but also effectively defend attacks, and the defense success rate is increased by 33%. This is of great significance for the safety protection of personal data in the car during the development of the auto-drive system, and will escort the research of personal data in the car during the subsequent development of the auto-drive system.

Keywords: Auto-drive system, personal data, SVM technology, hive mechanism

1. INTRODUCTION

With the rapid development of automatic driving technology, in the near future, a wide range of applications and services will be developed to operate the vehicle terminal, and personalized driving experience is highly sought after. But if we want to really achieve this goal, it is essential for the management and protection of user's personal data.

Many experts and scholars have conducted in-depth research on user personal data management under driverless system. Brell T said that the continuous development of autonomous vehicle (AVs) affects the future of transportation. In addition to the potential benefits in terms of safety, efficiency and comfort, it is also necessary to address the potential risks of new driving technologies. The research results help to understand the risk perception of automatic driving, which helps to promote the successful implementation of autonomous vehicle in the market and develop public information strategies [1]. Lu Hren said that due to the combination of artificial intelligence and the Internet of Things, automatic driving has attracted a lot of attention from academia and industry because of its economic and social benefits. However, a single autonomous vehicle with limited intelligence and the existing Internet of Vehicles architecture cannot guarantee ultra-low latency and ultra-high reliability. Based on the cloud/fog computing model and the artificial intelligence service framework of the Internet of Things, he proposed a cross-domain automatic driving solution [2]. Sun X firmly believes that trust is a major determinant of the acceptance of autonomous vehicle (AVs). Lack of appropriate trust may prevent drivers and the whole society from using this technology. He explored the effect of personalized autonomous vehicle as a new way for drivers to trust autonomous vehicle. The personalized autonomous vehicle system can recognize the user's driving behavior and accordingly adjust the driving style of autonomous vehicle [3]. Morales Alvarez W believes that in conditional automation (Level 3), human drivers can transfer driving dynamic tasks (DDT) to auto-drive system (ADS), and can only recover control in emergency situations, allowing them to participate in non driving related tasks (NDRT) when the vehicle is operated in its operating design field (ODD) [4]. They have made some contributions to the development of the driverless system, but they are not involved in the collection, utilization, protection and compliance governance of personal data in the vehicle during the development of the auto-drive system. Therefore, this paper will carry out further research.

*xiaoyineng@pku.edu.cn

Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023), edited by Srikanta Patnaik, Tao Shen, Proc. of SPIE Vol. 12779, 1277929 · © 2023 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2689722

Through sorting out the existing research and combining with comparative experiments, this paper conducts in-depth research on the collection, utilization, protection and compliance governance of personal data in the car in the development of the auto-drive system.

2. COLLECTION, UTILIZATION, PROTECTION AND COMPLIANCE GOVERNANCE METHODS OF PERSONAL DATA IN THE VEHICLE IN THE DEVELOPMENT OF AUTO-DRIVE SYSTEM

2.1 Auto-drive system

The so-called autonomous vehicle is an intelligent network driving process of driverless driving with the help of computer systems. The autonomous vehicle relies on the cooperation of artificial intelligence, visual computing, radar, monitoring equipment and global positioning system. The intelligent networked control system makes the vehicle and road perfectly fit without the intervention of the driver, and automatically and safely drives the motor car [5].

Automobile manufacturers and technology companies face huge obstacles when deploying autonomous vehicle, because safety standards formulated decades ago require manual control of auto-drive system. According to the old regulations of NHTSA, the vehicle "must have a driver's seat, a steering wheel and steering lever, or only one front outboard passenger seat". However, NHTSA stated that "vehicles completely controlled by the auto-drive system do not need manual driving control." Therefore, NHTSA amended the regulation.

NHTSA said that the current regulations do not prohibit the deployment of autonomous vehicle as long as they are equipped with manual driving control systems. NHTSA will continue to consider changing other safety standards, but manufacturers may still need to apply to NHTSA for exemption before they can sell their vehicles equipped with auto-drive system. In recent years, the penetration rate of the new energy vehicle market has been increasing under the condition of new vehicles, which has led to the entry of many industry giants. For example, the entry of Internet companies such as Xiaomi, Baidu and Huawei has surged again in this wave of intelligence. Admittedly, the birth of each brand must have its inherent ambition and blueprint ambition. The founders of each brand are eager to make their cars more colorful and even change the industrial structure and the whole world. To fully realize automatic driving, we need to solve technical problems, that is, effectively integrate multiple sensors in the entire vehicle equipment: millimeter wave, laser radar and camera, and analyze and process the sensor data through computer; Secondly, we need to solve data and map problems [6]. The map supporting unmanned driving is not two-dimensional abstract data, but requires high-resolution and high-precision data, so the amount of data stored and transmitted will be large. Once the data is connected to the system center of the vehicle, the intelligent automatic driving algorithm of AI technology can be used to identify the environment and realize intelligent automatic control [7-8].

As one of the few automobile driving companies in China that can achieve the mass production and delivery of TDA4VM, Nullmax has completed the fixed point of integration of parking lot and parking lot since 2020 and 2021. Nullmax has chosen the integration of parking and parking from the scene to the automatic driving landing from the early stage of its business, and has created the integration of parking and parking for the entire scene.

At present, the field of integrated navigation and berthing is divided into several forms:

The independent system of two SOC and two domain controllers is used for driving and parking, which is called 1.0 form;

Integrate two SOCs into a domain controller to reduce costs through a small amount of integration, which is called 2.0 form;

A single SOC and integrated domain controller are used, but the sensor still does not have the 3.0 form of deep reuse.

Automatic driving technology is not only the modernization of traditional cars, but also a major project to upgrade the entire driving system, involving the upgrading of the car itself, the upgrading and improvement of traffic infrastructure, the complement and establishment of social infrastructure, the maturity and wide application of intelligent network technology, the promotion of laws and regulations, and even the change of the driver's sense of responsibility and behavior habits [9]. As the upgrade of the whole transportation industry, automatic driving will change people's travel mode in the future. Its working principle mainly depends on the computer system inside the car and the sensors around

the car body, and the information is processed remotely through the cloud server and fed back to the vehicle to guide the driving, so that the car can drive itself without driver intervention [10], as shown in Figure 1.

Figure 1. Schematic diagram of the automatic driving system

According to the developed classification standard of the degree of automation, the degree of automatic driving of vehicles can be divided into several levels [11], as shown in Table 1.

Table 1. Degree of automatic driving of vehicles

	Eyes	Ears	Hand	Legs
LEVEL 0	√	√	√	√
LEVEL 1	√	√	or	
LEVEL 2	√	√		
LEVEL 3	√	√		
LEVEL 4	√	√		
LEVEL 5				

Automatic driving technology cannot be separated from the support of big data and algorithms. In this work, we studied the traditional autonomous driving target detection module using SVM (support vector machine) method in intelligent control algorithm. In the process of practical application, SVM needs to use kernel function for feature transformation [12].

The common classification of kernel function includes polynomial kernel function and Gaussian kernel function, and their corresponding forms are as follows:

Polynomial kernel function:

$$k(x, y) = (ax^T * y + c)^d \tag{1}$$

Where, a is the adjustment parameter and d is the maximum number of times.

Gaussian kernel function:

$$k(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}} \tag{2}$$

$\|\cdot\|$ represents the module of the vector.

2.2 Collection, utilization, protection and compliance management of personal data in the vehicle

2.2.1 Collection and utilization of personal data in the vehicle

According to the purpose of collecting data from driverless vehicles, data can be divided into two categories: traffic and lifestyle. Comprehensive access to external information is crucial to the safe driving of driverless vehicles. In the traffic category, the data generated using V2X technology can be further classified [13].

For personal data in the car, a large amount of data needs to be mined, processed and shared. For example, for the requesting end, we need to request service from the remote end and send the current vehicle status; For the server, we need to collect a large number of vehicle and road information, and then extract, analyze and contextualize the value data to provide personalized services for the requester in real time.

The process of extracting potentially useful knowledge or patterns from incomplete, fuzzy, noisy, highly random and previously unknown real-world application data is called data mining. If you use the traditional method to simply operate directly with text, you need to traverse the text line by line. The small amount of data may be faster, but the processing efficiency of massive data is not as high as Hive.

Hive is the basis of Hadoop-based data warehouse and is a simple and easy-to-use parallel data processing tool. In the big data environment of vehicle network, the limitations of traditional methods are increasingly prominent. Therefore, the data to be collected and extracted in this article will be imported using the Hive mechanism, as shown in Figure 2 [14].

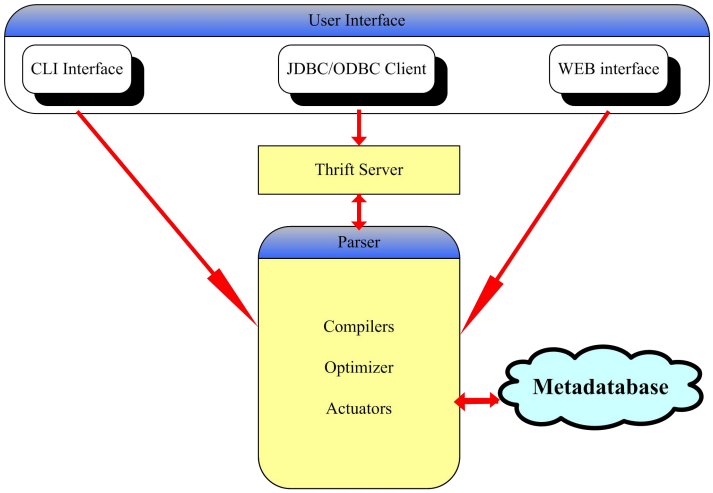

Figure 2. Hive architecture diagram

2.2.2 Protection of personal data in the vehicle

Users usually need to register and provide personal information to carry out activities on various Internet service platforms, while various behaviors and activity data are retained. The service platform collects personal data with the user's knowledge and consent, and many users do not know the use of personal data. Under the normal and legal transaction background, through the mining of user's personal data, the purchaser may locate the natural person through user's personal data identification.

Given the huge value of personal data, rational use of big data resources can promote social development and improve efficiency. However, in the context of legal big data, illegal acts such as stealing personal data and illegal trafficking also occur from time to time. Therefore, users themselves need to strengthen the awareness of security protection, and more importantly, the platform that accepts user personal information should be responsible for user data and protect the security of user data, which also brings great technical challenges to the data processing system of the platform.

2.2.3 Compliance governance of personal data in the vehicle

The value of data becomes self-evident with the arrival of the information economy era. More important than traffic data is the safety of location information and vehicle control information. Most of the data of networked vehicles are related to the location of vehicles, which is also closely related to people. If these data can be easily obtained, it is easy to track a person and find out their location, and the negative impact will be unimaginable. The personalized data of users can be identified and located, including all aspects of people's life. The protection of personal information is of great value for maintaining the personal dignity and freedom of the information subject, which is also an important reason for protecting the data of driverless vehicles.

Data compliance governance is no longer just to ensure the confidentiality and integrity of data itself, but also involves the rights and interests of individuals, enterprises, governments and other aspects, including many data governance issues. At present, the current data governance is becoming more and more mature. On the one hand, from the international perspective, the rule system is becoming increasingly complex; On the other hand, from the domestic situation, the management system has been gradually improved [15].

In order to strengthen the compliance governance of user's personal data, other efforts need to be made: at the legal level, domestic and foreign legislation should be strengthened; At the management level, we should improve the internal control mechanism of data security; At the technical level, data security protection needs continuous improvement.

3. EXPERIMENTAL ANALYSIS ON COLLECTION, UTILIZATION, PROTECTION AND COMPLIANCE GOVERNANCE OF PERSONAL DATA IN THE VEHICLE DURING THE DEVELOPMENT OF AUTO-DRIVE SYSTEM

This paper analyzes the personal data of 2000 different types of car owners in A city, and constructs an experiment from the speed of personal data collection in the car and the degree of personal data protection.

3.1 Comparison and analysis of personal data collection in the vehicle

This paper collects data through Hive, and obtains personal data in the car in two ways: Hive data collection and traditional collection. This paper collects 1000 data, of which 998 points are valid and 2 points are invalid. The selection criteria are based on the data from the last five years.

This study compares and analyzes the collection speed of Hive data collection and traditional collection of personal data in the car, and the results are shown in Figure 3.

Figure 3. Comparative analysis chart of in-vehicle personal data collection

From the data in Figure 3, we can see that the traditional data collection speed is: music preference 23 minutes; The temperature preference is 12 minutes; Use the light for 8 minutes; Place collection for 9 minutes; Common routes take 17 minutes. Hive data collection speed: 4 minutes for music preference; Temperature preference 1 minute; Use the light for 2 minutes; Place collection for 1 minute; Common routes take 3 minutes. It can be seen that Hive's data collection speed is far faster than the traditional way in any aspect, which will provide effective technical support for the research of personal data in the car in the subsequent development of auto-drive system.

3.2 Comparative analysis of personal data protection

This study compares and analyzes the number of attacks and the number of successful defense of the personal data system in the vehicle in the new and traditional modes within a week. The results are shown in Figure 4.

Figure 4. Comparative analysis chart of personal data protection

It can be seen from Figure 4 that the traditional data management system has been maliciously attacked 9 times and successfully defended 6 times in a week; The new data management system has been attacked twice and successfully defended twice at the same time. This shows that the use of a new data protection system can not only reduce the number of external attacks by about 33%, but also effectively defend attacks, with a 33% increase in the defense success rate. This is of great significance for the safety protection of personal data in the car during the development of the auto-drive system, and will escort the research of personal data in the car during the subsequent development of the auto-drive system.

4. CONCLUSION

In the era of big data, people enjoy the convenience brought by it, but also bear the corresponding risks. While people enjoy the extraordinary driving experience brought by the automatic driving system, the management of users' personal data is the problem that everyone is worried about. In this paper, Hive is used for data collection, and the data of Hive data collection and traditional collection are obtained. The experimental results show that Hive data collection speed is far faster than the traditional method in any aspect, which will provide effective technical support for the research of personal data in the car in the subsequent development of auto-drive system. This paper studies the collection, utilization, protection and compliance governance methods of personal data in the car in the development of the auto-drive system, aiming to improve the collection efficiency and utilization rate of personal data in the car through intelligent algorithms,

and protect the safety of user personal data and strengthen the compliance governance of data through a new data protection system. The development of driverless systems will continue to be valued, so the discussion of user personal data will not stop. Due to the limitation of academic level, the research of this paper is still shallow, and further research and discussion are needed in the future.

REFERENCES

[1] Brell, T., Philipsen, R., Ziefle, M., sCARy! Risk perceptions in autonomous driving: The influence of experience on perceived benefits and barriers[J]. Risk analysis, 39(2): 342-357 (2019).

[2] Lu, H., Liu, Q., Tian, D., et al., The cognitive internet of vehicles for autonomous driving[J]. IEEE Network, 33(3): 65-73 (2019).

[3] Sun, X., Li, J., Tang, P., et al., Exploring personalised autonomous vehicles to influence user trust[J]. Cognitive Computation, 12(6): 1170-1186 (2020).

[4] Morales-Alvarez, W., Sipele, O., Léberon, R., Tadjine, H. H., Olaverri-Monreal, C., Automated driving: A literature review of the take over request in conditional automation. Electronics, 9(12), 2087 (2020).

[5] Detjen, H., Faltaous, S., Pfleging, B., et al., How to increase automated vehicles' acceptance through in-vehicle interaction design: A review[J]. International Journal of Human–Computer Interaction, 37(4): 308-330 (2021).

[6] Hancock, P. A., Kajaks, T., Caird, J. K., et al., Challenges to human drivers in increasingly automated vehicles[J]. Human factors, 62(2): 310-328 (2020).

[7] Ulahannan, A., Thompson, S., Jennings, P., et al., Using glance behaviour to inform the design of adaptive HMI for partially automated vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 23(5): 4877-4892 (2021).

[8] Ren, K., Wang, Q., Wang, C., et al., The security of autonomous driving: Threats, defenses, and future directions[J]. Proceedings of the IEEE, 108(2): 357-372 (2019).

[9] Taylor, E., Autonomous vehicle decision-making algorithms and data-driven mobilities in networked transport systems[J]. Contemporary Readings in Law and Social Justice, 13(1): 9-19 (2021).

[10] Faisal, A., Kamruzzaman, M., Yigitcanlar, T., et al., Understanding autonomous vehicles[J]. Journal of transport and land use, 12(1): 45-72 (2019).

[11] Biondi, F., Alvarez, I., Jeong, K. A., Human–vehicle cooperation in automated driving: A multidisciplinary review and appraisal[J]. International Journal of Human–Computer Interaction, 35(11): 932-946 (2019).

[12] Devi, S., Malarvezhi, P., Dayana, R., et al., A comprehensive survey on autonomous driving cars: A perspective view[J]. Wireless Personal Communications, 114(3): 2121-2133 (2020).

[13] Van Wyk, F., Wang, Y., Khojandi, A., et al., Real-time sensor anomaly detection and identification in automated vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 21(3): 1264-1276 (2019).

[14] Alsulami, A. A., Abu Al-Haija, Q., Alqahtani, A., et al., Symmetrical Simulation Scheme for Anomaly Detection in Autonomous Vehicles Based on LSTM Model[J]. Symmetry, 14(7): 1450 (2022).

[15] Jameel, F., Chang, Z., Huang, J., et al., Internet of autonomous vehicles: architecture, features, and socio-technological challenges[J]. IEEE Wireless Communications, 26(4): 21-29 (2019).