Collecting (Personal) Passenger Data in Public Transport or Do Carriers Really Need Our (Personal) Data? – An Overview of the Situation in the Republic Of Croatia

Andrej Ignjatić*, Goran Vojković**
* University of Zagreb, Faculty of Organization and Informatics, Varaždin, Croatia
** University North, Koprivnica, Croatia
anignjatic@student.foi.hr, gvojkovic@unin.hr

2024 47th MIPRO ICT and Electronics Convention (MIPRO) | 979-8-3503-8250-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/MIPRO60963.2024.10569751

***Abstract* — This paper investigates the volume, or the amount of (personal) data that a user (passenger) must give to the public transport organizations. The goal of the research is to structure and present, using the example of the largest cities in the Republic of Croatia, which (personal) data public transport organizations can access on their mobile devices, and to answer the research question: What (personal) data do public transport organizations collect within the Republic of Croatia? During the research, the Google Store was accessed, applications used for fare collection were checked, and data that the user must accept in order to use the transport service within public transport, and to be able to pay for that service with a mobile device, were structured (in table). Based on this data, a model proposal was made that shows the data collected and processed, used for fare collection.**

Keywords - (personal) data; public passenger transport; legislation; fare collection; applications

I. INTRODUCTION

Public passenger transport companies (public transport organizations), in accordance with technological development, are improving ways to fare collect for their services, starting from paper tickets, through smart cards, to mobile devices. Using mobile devices for fare collection can be seen as one segment of smart cities.

By installing and using different applications, users give various permissions to organizations, which include access to personal data. The legitimate interest of organizations in accessing such data is questionable. Research question which this paper is trying to answer is: What (personal) data do public transport organizations collect within the Republic of Croatia?

This paper presents an overview of permissions that a user must allow organizations in order to use their services, including, but not limited to, fare collection. This is the first paper that shows structured overview of the permissions that a user must accept in order to use the services of public transport organizations in the four largest Croatian cities, and to be able to pay for the used service through public transport.

The rest of the paper is divided as follows. The second section presents basic definitions related to the subject of the work and presents relevant research. The third section presents the research methodology. The fourth section presents the results of the research, which are discussed in the fifth section. At the end of the paper, there is a conclusion given.

II. BASIC DEFINITIONS AND RELEVANTN RESEARCH

There is no agreed definition of a smart city. Thus, in [1] it is stated that there are two groups of researchers. One group of researchers defines a smart city as a means of sustainable urbanization. The other group of researchers defines a smart city as technological progress reflected in everyday life.

A city can be considered smart if information and communication technologies are used along with the Internet of Things to connect certain or all functional systems of a city [2], [3], in order to raise the level of quality of life, competitiveness, and operational effectiveness of urban systems [4]. It is then necessary to understand how the Internet of Things generates a large amount of data, especially during the development of the service [5], [6], [7], [8], and to distinguish the need for collecting, accessing, and managing data [2], which is often processed by artificial intelligence [9], [10].

With smart cities and the Internet of Things, data must be collected by certain devices. For this purpose, various sensors are most often used, for the purpose of collecting data such as: GPS location, steps, pulse, and others. Lately, mobile devices and watches are often used for collecting data, considering that these devices have become a basic part of everyday life [11], (in [11] citing 1 and 3).

In the process of passenger transport, the purpose of introducing smart city concepts, which in this case would represent intelligent transportation systems (ITS), can be to provide better service to passengers (in [8] citing 7).

The primary legal framework that regulates the protection of personal data within the European Union is Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, while revoking Directive 95/46/EC [12]. This regulation is globally known under the acronym GDPR (further in the text: 'General

Regulation'). The purpose of the General Regulation is to protect privacy and protect against potential data breaches for all citizens of the European Union in a data-driven world. Additionally, the General Regulation defines personal data as any data on the basis of which an individual can be identified or serves as an aid for identifying an individual [2].

In Recital 4 of the General Regulation, its general goal is stated. Namely, it is stated here that "The right to the protection of personal data is not an absolute right; it must be considered in relation to its function in society and be balanced against other fundamental rights, in accordance with the principle of proportionality" [13].

In addition to the General Regulation, another regulation has been introduced in the protection of citizens' data within the Union - Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act) [14]. This regulation focuses on data generated from Internet of Things solutions and relevant services, while excluding large technical companies, digital platforms, and telecommunications operators, which are considered to be the largest manufacturers and owners of data [15]. The General Regulation directly applies in the Republic of Croatia, but parts that need to be regulated by national legislation are resolved by the Law on the Implementation of the General Data Protection Regulation [16].

The Republic of Croatia has adopted the Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 concerning measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 [17] into its national legislation. Thus, Directive 2022/2555 was transposed into the Cyber Security Act [18]. Additionally, it is important to mention Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act) [19], which serves several purposes. First, this Regulation has a key role in enhancing cyber security within European Union. Second, it establishes a legal framework for certification of information and communication technology security. And third, this Regulation expands the powers and role of ENISA, the European Union Agency for Cybersecurity.

The method of automated processing is defined in the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data [20]. Article 5 of the Convention defines that personal data which are the subject of automated processing should be [13]:

- Collected and processed in good faith and lawfully;
- Stored for specific and lawful purposes and must not be used in a manner incompatible with those purposes;
- Adequate, relevant and not excessive in relation to the purposes for which they are stored;
- Accurate, and if necessary, updated;
- Preserved in a form that allows identification of data subjects for a period not longer than that required for the purposes for which they are stored.

III. METHODOLOGY

Initially, the Google Store website [21] was accessed on June 11, 2023, and applications used for fare collection via mobile devices in Zagreb, Split, Rijeka, and Osijek were sought. The search was performed based on the term of the name of the passenger transport organization in the mentioned city. Additionally, an application for fare collection in national rail transport carrier was analysed.

The reference applications analysed are those published by the public transport service organization, except in the case of one application. More about this in the Results section.

Regarding the data collection two steps were taken. Firstly, all of the applications were accessed via a personal computer (Windows operating system). On the personal computer, access was made to the part of the website that defines access rights and protection methods, where relevant data were collected. Secondly, to check given permission on Android Smartphone application called Exodus [22] was installed. However, APP-4 was not analysed with Exodus because that organization later changed the method of fare collection and withdrew that application. None of the applications' functionalities were checked on the mobile phone. While launching all the applications, all requested permission were granted.

IV. PRESENTATION OF COLLECTED DATA

Applications for two cities were identified as defined in the Methodology section. For one city, an application was found for which the publisher is one network service operator within the Republic of Croatia. For one city, no such application was found.

Based on the access to the section of the website that defines access rights and protection methods, relevant data were extracted. Such data are presented in Table 1. The applications, or organizations, are not explicitly mentioned, but the applications are coded with labels from APP-1 to APP-4, where the label APP represents the acronym for the term application, and the labels 1-4 represent the order of the analysed application.

The following is an explanation of the legend of Table 1. For cells marked with: ✓ represent the label "yes", while empty cells represent the label "no". In the case where a cell contains the mark: M, then it is mandatory, and in the case of the mark: O, then it is optional. The letter mark in the cell indicates the reason for the obligation or optionality, so the mark [a] represents App functionality, [b] Analytics, [c] Fraud prevention, [d] Security, and compliance, [e] Account management and [f] Personalization.

The rows in Table 1. represent the cumulative data collection points for all permissions, regardless of the application. Specifically, if the application does not collect or request access to a specific type of data, the cell will remain empty.

TABLE I. PRESENTATION OF THE OBTAINED RESEARCH RESULTS

	APP-1	APP-2	APP-3	APP-4
Functionality of purchasing				
Individual transportation tickets	✓	✓	✓	✓
Monthly and annual passes	✓	✓	✓	
Access and data collection				
Name		**O** [a, e, f]		**M** [a-e]
Location	**M** [a]			
Address				**M** [a-e]
Phone number				**M** [a-e]
Email address	**M** [a, e]	**M** [e, f]		**M** [a-e]
User IDs		**M** [a, e]		**M** [a-e]
Financial info (purchase history)	**M** [e]			**M** [e]
Messages (1 - e-mail, 2 - within APP)	**1: M** [a, f]	**2: O** [a]		**1: M** [a-e]
Other identifiers (1 - device or other, 2 - not specified)		**1: M** [a, e, f]		**2: M** [a-e]
Data protection				
Sharing data with third parties				
Data encryption during transmission	✓	✓	no transmission	✓

V. ANALYSIS OF COLLECTED DATA

Table 1. is discussed. All four applications have the same functionalities - purchasing tickets for individual transport. Only one application (APP-4) did not have the functionality of selling monthly and annual passes, while the remaining three have this functionality.

It is necessary to discern how it is necessary to achieve a balance, which needs to be thoroughly considered as the collection of additional personal data can improve the quality of services, but can also lead to a user's privacy violation [23], and achieving the mentioned balance is very important and presents a challenge [24]. According to the collected results, it can be seen that two applications (APP-1 and APP-4) are very invasive if we consider the type of data collected. It is particularly necessary to note that the application APP-1 requires the use of location, i.e., tracking the position of the user's device, for the general functioning of the application, i.e., purchasing a transport ticket. Also, it is necessary to examine the application APP-4, which is even more invasive than the application APP-1. This application requires access to a large amount of data, including but not limited to: full name, address, and mobile device number. Furthermore, it is important to mention that APP-1 is the only application that did not indicate the ability to audio record, although according to the Exodus, audio recording was activated for the use of APP-1.

Considering the legitimacy of personal data processing also applies if processing is necessary for the performance of a contract to which the participant is a party or to take actions at the request of the participant before entering into a contract [13], it is necessary to discern whether the amount of data collected for public transport service fare collection is really necessary for the fulfilment of that contract. If we observe that some of the collected data are used for designing routes and schedules of public transport, in order to attract more passengers [25], or for optimization [8], and the data are indeed anonymized, then such data collection can be considered legitimate. Indeed, if person X travels from location A to location B at time T, it is not necessary for the organization to fulfil the contract. However, for the organization, the aggregated data about the existence of people traveling from location A to location B at time T is a relevant data, because of the potential need for optimization of the business (transport) process.

European Union has anticipated the possibility of collecting a large amount of data (in regard to artificial intelligence). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [26] defines that data that are not biometric user's data can be collected, provided that such collected data do not constitute personal data and do not belong to special category of personal data.

It is also important to note that there are two applications that are not overly invasive (APP-2 and APP-3). Here it is important to thoroughly review the application APP-3 which does not collect any data about users and there is no data transfer to the organization. However, while using Exodus, it was observed that upon reviewing the data from Table 1, some permissions were found that users grant when using the applications. For example, APP-2 and APP-3 access the user's location.

Regarding the data collection via Internet of Things devices, Vojković and Milenković in [27] states that utility providers should inform users what is the purpose of certain measurements, i.e. collection of certain data. The same can be observed in the context of public transport. Although, it is stated in the Google Store why certain data is collected, it is necessary to better explain this to service users especially considering which data is collected. In this regard, one has to question whether organizations really

need this data for providing service, i.e., fare collecting. Furthermore, some users may find the reasons for which access is requested to certain data disputable and therefore refuse to give access.

Although Bu-Pasha in [7] believes that the best foundation for collecting and processing personal data is voluntary consent, one has to question whether not giving voluntary consent can be a basis for denying service? In fact, as it is mentioned in the previous paragraph, due to access to a large amount of data, certain users may not consent to collection of some data. Consequently, does this imply the possibility of denying the right of access to services. We argue that denial should not be allowed, but access to various fare collection options should be permitted, including paper tickets, smart cards, and others. Also, transport organizations always have the possibility of using other (non-invasive) technologies for monitoring traffic demand, including but not limited to various sensors.

Given the above, one of the ways of future development of applications could be based on the potential limitation of (personal) data that the transport organizations can collect. In fact, access to a large amount of personal data is not necessary for providing transportation services and for fare collection. Furthermore, even under the assumption that such access is necessary, personal data can be used for the longest time needed for the service itself [12]. Therefore, the proposal suggests creating a model that would define the data that organizations can collect and process for providing transportation payment services and under which conditions, as shown in Table 2.

Every user within the ITS system must be authenticated [28]. In order to ensure the property of user authentication, it is suggested to use tokens as user identifiers. In this way, the user cannot be connected with other personal data. By following tokens, the service provider can recognize the traffic demand and thus provide better service to users, and the user can know that he or she cannot be associated with a certain token.

VI. CONCLUSION

This paper investigated the amount of data, or rather, the volume of data to which public transport organizations require access, so that users can enjoy the transport service, i.e. to be able to pay transport tickets via mobile devices. The research was conducted in the four largest cities in the Republic of Croatia. One city does not have application for transport payment via mobile device on the Google Store website. The collected data is structured in Table 1.

The research noted that one application is extremely invasive when accessing and collecting (personal) data of users. Additionally, the research found that none of analysed applications fully discloses all data collection methods to the user. For instance, APP-1 does not specify that it collects audio data, while APP-2 and APP-3 fail to disclose that they gather users' location data.

Based on the collected data, a standardized model for collection and processing of data, which can then be processed, for transport payment was proposed. The proposed model is minimally invasive to users and users have autonomy when choosing to give the right of access to data to the service provider.

For conducting further research, it is suggested to compare the data that can be collected manually, for example if a user buys a paper ticket, or pays transport with a smart card, compared to data collected via mobile devices, as well as comparison of ways of processing differently collected data. Also, it is suggested to compare the general terms of use of each application and the General Regulation.

TABLE II. THE PROPOSED DATA COLLECTION MODEL

	Proposal	Explanation
Access and data collection		
Name	No	
Location	No	
Address	No	
Phone number	No	
Email address	Optional	Optionally, the user can choose whether to allow access to the address, in order to receive informational messages about the payment.
User IDs	Mandatory	The user identifier represents a token that cannot uniquely determine the user. By scanning a two-dimensional code when entering and exiting, the user indicates that he or she is in a certain vehicle and thus signals the organization about the traffic demand.
Financial info (purchase history)	Optional	The user can choose whether to save financial data, which would represent previous transport payments.
Messages	Optional	The user can choose whether to receive informational messages about transport payment via e-mail or within the application.
Other identifiers	No	
Data protection		
Sharing data with third parties	No	Data is stored locally so there is no data transfer and the same cannot be shared.
Data encryption during transmission	No	

REFERENCES

[1] A. Founoun, A. Hayar, and A. Haqiq, "The Textual Data Analysis Approach to Assist the Diagnosis of Smart Cities Initiatives," in *2019 IEEE International Smart Cities Conference (ISC2)*, Oct. 2019, pp. 150–153. doi: 10.1109/ISC246665.2019.9071663.

[2] G. Vojkovic, "Will the GDPR slow down development of smart cities?" in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 2018, pp. 1295–1297. doi: 10.23919/MIPRO.2018.8400234.

[3] M. Srebalová and T. Peráček, "Effective Public Administration as a Tool for Building Smart Cities: The Experience of the Slovak Republic," *Laws*, vol. 11, no. 5, Art. no. 5, Oct. 2022, doi: 10.3390/laws11050067.

[4] L. Li, A. Taeihagh, and S. Y. Tan, "What factors drive policy transfer in smart city development? Insights from a Delphi study," *Sustainable Cities and Society*, vol. 84, no. 104008, Sep. 2022, doi: 10.1016/j.scs.2022.104008.

[5] H. M. K. K. M. B. Herath and M. Mittal, "Adoption of artificial intelligence in smart cities: A comprehensive review," *International Journal of Information Management Data Insights*, vol. 2, no. 100076, Apr. 2022, doi: 10.1016/j.jjimei.2022.100076.

[6] R. Juvenile Ehwi, H. Holmes, S. Maslova, and G. Burgess, "The ethical underpinnings of Smart City governance: Decision-making in the Smart Cambridge programme, UK," *Urban Studies*, vol. 59, no. 14, pp. 2968–2984, Nov. 2022, doi: 10.1177/00420980211064983.

[7] S. Bu-Pasha, "Legal aspects, public interest, and legitimate interest in processing personal data to operate autonomous buses in the regular transportation system," *SECURITY AND PRIVACY*, vol. 5, no. 5, 2022, doi: 10.1002/spy2.247.

[8] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big Data Analytics in Intelligent Transportation Systems: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 1, pp. 383–398, Jan. 2019, doi: 10.1109/TITS.2018.2815678.

[9] O. Golubchikov and M. J. Thornbush, "Smart Cities as Hybrid Spaces of Governance: Beyond the Hard/Soft Dichotomy in Cyber-Urbanization," *Sustainability*, vol. 14, no. 16, 2022, doi: 10.3390/su141610080.

[10] D. Luckey, H. Fritz, D. Legatiuk, K. Dragos, and K. Smarsly, "Artificial Intelligence Techniques for Smart City Applications," in *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering*, E. Toledo Santos and S. Scheer, Eds., Cham: Springer International Publishing, 2021, pp. 3–15. doi: 10.1007/978-3-030-51295-8_1.

[11] T. S. Pias, D. Eisenberg, and M. A. Islam, "Vehicle Recognition Via Sensor Data From Smart Devices," in *2019 IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE)*, Oct. 2019, pp. 96–99. doi: 10.1109/ECICE47484.2019.8942799.

[12] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)." Accessed: Jun. 10, 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/en/TXT/HTML/?uri=CELEX%3A32016R0679

[13] G. Vojković, *Medijsko pravo*. Koprivnica: Sveučilište Sjever, 2023.

[14] "Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)." Accessed: Jun. 10, 2023. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020PC0767&from=EN

[15] I. Calzada, "Citizens' Data Privacy in China: The State of the Art of the Personal Information Protection Law (PIPL)," *Smart Cities*, vol. 5, no. 3, Art. no. 3, Sep. 2022, doi: 10.3390/smartcities5030057.

[16] "Zakon o provedbi Opće uredbe o zaštiti podataka. NN 42/18." Accessed: Jun. 11, 2023. [Online]. Available: https://narodne-novine.nn.hr/clanci/sluzbeni/2018_05_42_805.html

[17] "DIRECTIVE (EU) 2022/2555 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive)." Accessed: Apr. 14, 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX%3A02022L2555-20221227

[18] "Zakon o kibernetičkoj sigurnosti. NN 14/24." Accessed: Apr. 14, 2024. [Online]. Available: https://narodne-novine.nn.hr/clanci/sluzbeni/2024_02_14_254.html

[19] "Regulation - 2019/881 - EN - EUR-Lex." Accessed: Apr. 14, 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2019/881/oj

[20] "Directive - 2016/1148 - EN - EUR-Lex." Accessed: Apr. 14, 2024. [Online]. Available: https://eur-lex.europa.eu/eli/dir/2016/1148/oj/eng

[21] "Android Apps on Google Play." Accessed: Jun. 11, 2024. [Online]. Available: https://play.google.com/store/games?hl=en_US&gl=HR

[22] Exodus Privacy, "Exodus." Accessed: Apr. 14, 2024. [Online]. Available: https://play.google.com/store/apps/details?id=org.eu.exodus_privacy.exodusprivacy&hl=en_US&gl=HR

[23] C. Kalloniatis, D. Kavroudakis, A. Polidoropoulou, and S. Gritzalis, "Designing Privacy-Aware Intelligent Transport Systems: A Roadmap for Identifying the Major Privacy Concepts," *International Journal of Applied Geospatial Research*, vol. 10, pp. 73–91, Nov. 2018, doi: 10.4018/IJAGR.2019010104.

[24] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, N. Kumar, and M. M. Hassan, "A Privacy-Preserving-Based Secure Framework Using Blockchain-Enabled Deep-Learning in Cooperative Intelligent Transport System," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 9, pp. 16492–16503, Sep. 2022, doi: 10.1109/TITS.2021.3098636.

[25] G. Wu, Y. Ding, Y. Li, J. Luo, F. Zhang, and J. Fu, "Data-driven inverse learning of passenger preferences in urban public transits," *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, pp. 5068–5073, Dec. 2017, doi: 10.1109/CDC.2017.8264410.

[26] *Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. 2021. Accessed: Apr. 14, 2024. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

[27] G. Vojković and M. Milenković, "IoT Devices and the Need to Inform Utility Users of Collecting, Controlling and Processing of Personal Data," in *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, Sep. 2020, pp. 1470–1475. doi: 10.23919/MIPRO48935.2020.9245287.

[28] Q. E. Ali, N. Ahmad, A. H. Malik, G. Ali, and W. U. Rehman, "Issues, Challenges, and Research Opportunities in Intelligent Transport System for Security and Privacy," *Applied Sciences*, vol. 8, no. 10, Art. no. 10, Oct. 2018, doi: 10.3390/app8101964.