Privacy concerns on the mobility of smart cities

Tatiane Borchers[1], Victor Garcia Figueirôa-Ferreira[1], Ricardo Augusto Souza Fernandes[1]
Federal University of São Carlos (UFSCar), São Carlos, Brazil
tatiane@estudante.ufscar.br
victor.figueiroa@estudante.ufscar.br
ricardo.asf@ufscar.br

***Abstract* - The smart city concept arises in the last decades to improve quality of life. One characteristic of smart cities' strategies is smart mobility, which uses Information and Communication Technologies and Internet of Things for better traffic management and provision of new mobility services. This study aims to shed light on the discussion about data privacy, once this process will increase the availability of user-generated and user-centered information regarding urban mobility. A review was conducted to list the data collected and the major threats regarding data privacy in smart mobility initiatives. Results show that most of the data collected is highly sensitive and the major threats are related to the identification and tracking movements of users, besides leakage and unauthorized use of data. Thus, privacy needs to be strongly addressed in technological and regulatory developments to protect users' information.**

Index terms* - *Data privacy; Data protection; Smart cities; Smart mobility; Surveillance technologies.

I. INTRODUCTION

The concept and strategy of smart cities development is understood as a model in which there is a high investment in Information and Communication Technologies (ICTs) and Internet of Things (IoT), as well as in human and social capital, in order to promote quality of life in a universal way, *i.e.*, the smart city is multidimensional [1]. In this sense, with the consolidation of smart city research, several related topics have also been studied. Giffinger et al. [2] present six characteristics of a smart city: 1) smart economy or the level of competitiveness; 2) smart people, the social and human capital; 3) smart governance, with emphasis on the participation of society; 4) smart mobility, in general the use of ICT's and IoT in urban transportation; 5) smart environment, the management and use of natural resources; and 6) smart life, *i.e.*, quality of life. Therefore, smart cities initiatives involve a variety of components, including ubiquitous sensing devices, large-scale databases, and powerful data centers to collect, store, and intelligently analyze real-time information. Despite the potential of smart projects, security to protect massive data and privacy concerns remains to be carefully addressed [3], [4].

In terms of smart cities mobility, it is expected a shift towards what is called mobility as a service (MaaS), where the ownership of vehicles is replaced by usership, so instead of purchasing a vehicle, individuals will purchase the right to access mobility services provided by others, mainly corporations. This process will increase availability of user-generated and user-centered information. It is also expected to have an increase of intelligent infrastructure, electrification of fleets and automated vehicles, requiring processing of big data to match demand provision in real-time and to optimize systems' performance [3].

While open (public) data is useful for boosting data-driven intelligence to cities, data that is sensitive and not publicly available is often essential to understanding city challenges and needs. There is then a paradigm in planning that pervades the privacy and confidentiality of such data, mainly regarding the increasing collection, processing, and dissemination of people's private lives [5], [6].

This paper seeks to bring up the discussion of urban mobility data collection in the context of smart cities, aiming to present the main data collected and the related privacy threats. The remainder of this study is structured into a background review, presentation of the adopted methodology and the results, followed by drawn conclusions.

II. BACKGROUND REVIEW

A. *The historical construct of the smart city*

In the 1950s, in an interview, Albert Einstein stated that three great bombs, the demographic, the atomic and telecommunications, would mark the twentieth century. These would cause a flood of information, because of the exponential, explosive and chaotic nature of their development [7]. Bringing this statement to the reality of ICTs, the scenario is the same, but the flood is even bigger. The raw quantity of data multiplies every instant, as well as the density of connections between them. In the midst of this flood of information and technologies to support, manage, access and police data is the urban environment, the cities, the original hubs that now exist on two planes, the real and the virtual [7]-[9].

In this scenario, a set of strategies and models of management and development of the city, the urban environment and its inhabitants emerge, based on the use of this information through new technologies. After years of disagreement and debates about what a smart city actually was, in the early 2010s, some agreement began to exist.

Over the last decade, modern technology developments directed the cities to intelligently optimize the scarce resources and provide pervasive resources for all citizens. The outcomes of the smart cities have motivated researchers to focus on promoting and developing new smart solutions, which architecture can be summarized in four layers: (i) sensing, with different components and instruments to collect data from surrounding environment; (ii) data collection and

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001.

storage from different sources, such as homes, traffic, and citizens; (iii) data processing, where preprocessing techniques are performed; and (iv) smart processing and application, responsible for exchanging data between operators and smart applications [4].

B. *Smart mobility systems*

As previously mentioned, the aspect of mobility of a smart city presents, between other components, the existence of intelligent infrastructure, electrification of fleets, automated vehicles and a massive orientation to user-generated and user-centered information [3]. Smart mobility also collects real-time traffic information through sensors in vehicles or mobile applications used by drivers to provide support for traffic monitoring and management [10]. In this sense, some of the key strategies regarding smart mobility are: (i) automated vehicles; (ii) ride-sourcing; (iii) bike sharing; and (iv) initiatives on public transport.

Automated vehicles can make decisions independently of human and rely on sensor data and artificial intelligence to interpret the data, making decisions related to vehicle operation and adapting to changing conditions. These vehicles can exchange information through vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication [11]. Ride-sourcing is an arrangement where a driver provides a car and the driving service, and a passenger requests a ride through a third-party application. As one strategy of MaaS, its aim is to decrease car ownership and increase usership [3], [12]. Bike sharing is of widespread use in smart city environments and consists of a public or private service in which bicycles are made available for shared use to individuals on a short-term basis [13]. Finally, in public transport systems, fare payments through cards provide passengers flows and GPS (Global Positioning System) monitoring provides a real-time management over the network [14]. In addition, passengers’ flow and trajectories collected by companies are shared with internal and external agencies for trajectory mining and traffic management [15].

C. *Data privacy*

Information privacy was defined by Westin [16] as the right to select what personal information of an individual can be used for another person/company. Privacy protection is of high concern in any personal-data-related services, and in the case of smart cities preserving data privacy relies on acquiring user consent, assuring transparency, compliance, reliability and anonymity in all layers of smart cities’ architecture [4], [17], [18].

Ziegeldorf et al. [6] proposed a privacy definition related to IoT where is guaranteed to the subject awareness of privacy risks imposed by smart things and services, individual control over the collection and processing of personal information, awareness and control of subsequent use and dissemination of personal information to any entity outside the subject’s personal control sphere. The definition seeks to provide informational self-determination to subjects.

In the regulation sphere, the European General Data Protection Regulation (GDPR) [19] considered by many, despite severe criticism, the most ambitious movement to control privacy and personal data, mainly because it requires data controllers to obtain explicit consent for the processing of personal data from data subjects. Smart city technologies already exist and are extensively used, which does not mean that the struggle and arguments against them are not valid and do not deserve to be discussed. Initiatives such as the GDPR should be expanded and enhanced to return to individuals, as far as possible, control and disposition over their own data. Democratic, transparent and participatory control should be an essential element of future public policies and laws addressing these issues.

III. Methodology

The methodology propose in this paper is structured in three stages: (i) collection of bibliographic data and bibliometrics analysis; (ii) analysis of the type of data collected regarding mobility in smart cities; (iii) review of potential threats related to such data.

In stage 1, a search was performed in Scopus database on August 24, 2021, using the following query: ((“smart city” OR “smart cities” OR “big data”) AND (“mobility” OR “transport*”) AND (“data privacy” OR “data protection”)).

The search was refined by language and document type, being kept only articles published in journals and in English. In the first part of the search, the terms “smart city”, “smart cities” and “big data” are used to ensure that the articles returned were related to smart initiatives. The results obtained were downloaded and processed in Excel® and VOSViewer [20], [21] to build and visualize the bibliometric networks.

In stages 2 and 3, the articles returned in the search were analyzed individually to list the main data that are collected in the context of mobility in smart cities and the data privacy concerns associated with them.

IV. Results

The search resulted in 80 articles, with 326 authors and a set of 329 related keywords. These results are detailed in the following subsections, as well as the discussion of data collection and privacy threats.

A. *Bibliometrics analysis*

As seen in Fig. 1, the rising of research in this area is recent, with less than a decade of development. Even though it is a recent theme, with few publications since 2013, it is possible to see an increasing trend in publications, which peak occurred in 2020 with 24 publications. This pattern was expected because the smart cities’ researches themselves are recent and the privacy concerns were expected to grow with the modern technology developments over the last decade.

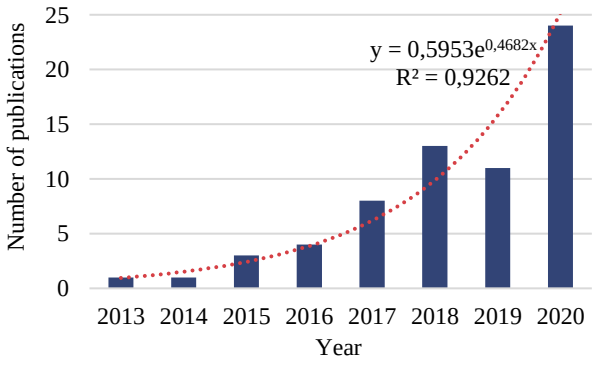

Fig. 1. Historical evolution of publications indexed in the Scopus database.

The countries with the most publications are respectively United States (20), China (17), Canada (7), Saudi Arabia (7), Australia (5), Germany (5), India (5) and Singapore (5). The most recurring keywords are the ones used in the search and

the ones that represent generic terms, as presented in Fig. 2. The keyword map show clusters related to generic words such as privacy, security, big data, internet of things and smart city are older and more connected, while from 2021 new clusters arise, representing emerging technologies or applications, such as 5G and data fusion.

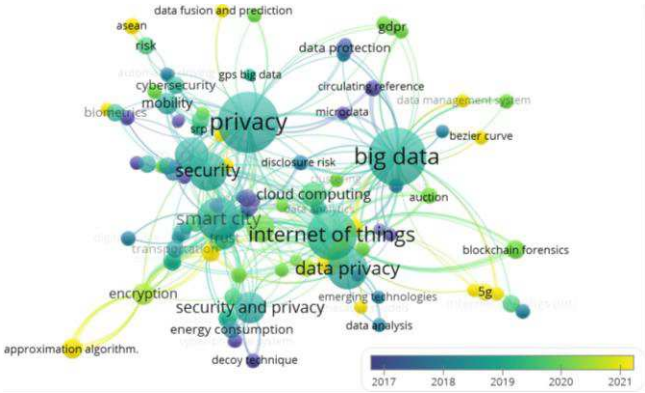

Fig. 2. Keywords map for the search with the terms "smart city", "smart cities", "big data", "mobility", "transport*", "data privacy" and "data protection" by year of publication.

B. *Data collected in mobility of smart cities*

Connected environment expects a user to transmit and share sensed and user generated data [22]. Personal information is collected by services available in smart cities and includes many attributes: (i) key identifiers, such as name, ID-number, and social security number; (ii) quasi-identifiers, as zip code, age, gender, birth-date, and address; (iii) sensitive data as salary and health condition; and (iv) other attributes that represent no danger if disclosed to untrusted third parties [23].

Another type of personal information collected is biometrics, which are mostly applied to control access to a place or service and represent human physical or behavioral characteristics that allow distinguishing two different people. In most cases, it is advantageous when compared to other identity certifications, because biometric characteristics can give a high percentage of guarantee to the authenticity of the claimant. DNA, iris, facial and hand geometry, fingerprint, palm-print are examples of biometric data [24].

Location-aware devices and systems capable of collecting localization and tracking users have been implemented in many smart city infrastructures over the last years [15]. GPS on vehicles, mobile phones, roadside unit, connected vehicles and urban traffic cameras can track human mobility on a massive scale, allowing urban planning and traffic management improvements towards a smarter mobility [10], [25], [26].

In public transportation, operators collect fees via magnetic tickets enabling them to know which user entered the system and at what time, generating big data of the movement of passengers. No additional data collection is necessary to obtain detailed information on the flow of passengers, tabulating and analyzing the electronic records is sufficient for better management of the system [14], [15], [27].

Regardless of the type of information, the amount of data generated is massive. This big data has a high potential to improve mobility statistics and researches on transportation issues, but despite its high potential there are certain problems related to their practical use. Firstly, big data is composed mostly of proprietary information belonging to private companies, with no obligation to share such information with the public, on the contrary, many companies have their estimated value directly linked to their user's database. Then, accessing this information is difficult for anyone outside such companies, including urban planners. This raises a question about these projects: Are they aimed towards a smart or only corporate cities? Secondly, a particular set of big data does not necessarily statistically represent the entire population. Lastly, there is a key issue regarding privacy. Using big data for research or urban planning will require information management rules for strictly protecting privacy [27].

C. *Threats*

Privacy preservation consists in protecting the identity of the user from outsider intruders and from insider breaches, avoiding unauthorized usage [28]. The advancement of technology and regulation must provide effective privacy protection or it might lay the groundwork infrastructure for a surveillance society [29]. In this subsection, the major threats regarding data privacy are presented. It is worth to point out that privacy concerns range from individual to community level, because data generated from the smart cities can reveal unforeseen information, which may not be evident otherwise [22].

The citizens' data is usually accessed by different companies, government agencies, and business partners for different purposes [4]. Attacks to identify users or compromise their personal information are one of the most critical concerns related to privacy [17], [22]. Identification denotes the threat of associating a key identifier, or a pseudonym of any kind, with an individual [6]. Pseudomyms or pseudo identifiers are a strategy to anonymize users' identity, being replaced frequently for further protection. Its application can occur with any key identifier of a person or even with vehicles that can send messages over the wireless channel of a vehicular network [15], [17], [29], [30]. However, the uniqueness of human spatiotemporal mobility trajectory can lead to user identification [11], [17]. In vehicular networks, for example, even if drivers can not be immediately identified, they can be re-identified because malicious actors can track their movements by picking up drivers' signed wireless broadcasts at different locations. Besides that, each pseudonym can still be resolved to a base identity by the authority that signed it [25], [29]. Eckhoff and Sommer [29] argue that complete privacy cannot be achieved, but they reinforce that it is time to place a stronger emphasis on privacy, adopting measures for the technical protection of users' information.

Concerning biometric data and its technologies, it can be observed a trend to generate more information than is needed for a specific purpose, *i.e.*, collection of other data less sensitive than biometrics could be employed but biometrics are used anyway, and this extra information can be used for purposes that were not legally allowed [31]. Besides that, its indiscriminate use and storage by many institutions may let this sensitive data susceptible to attacks. Biometrics characteristics are unique and invariable over time, and a leak could compromise data forever [24]. Moreover, its uniqueness can be used for surveillance as in the example presented by Icasiano and Taeihagh [12] where in 2020 the privacy commission of Philippines had to prohibit Grab (ride-sourcing company) from monitoring its riders with video and audio equipment.

Concerning localization and tracking, privacy refers to guarantee and preserve protection over the physical location of the user. This is a threat that has a strong link with identification because, although key identifiers are anonymized, sequences of locations can reveal a user’s frequently visited positions and preferred routes, allowing identification and sometimes bringing life-threatening consequences [25], [28]. As location-aware devices are employed by different services, user data is vulnerable in many datasets.

In ride-sourcing services, location-aware data might include trip data, location data, and other data that would make users and their activities easily identifiable [12]. In vehicular networks, because of accumulation of user data (as detailed location information) and its decentralized structure, much of the information generated is broadcasted wirelessly, allowing exploitation of this information for profit or surveillance by operators, other drivers, or arbitrary people [25], [29]. Other attributes related to locations, for example speed, may reveal user’s driving habits [25].

As presented by Hasan et al. [13], bike sharing data generally is publicized with user’s visited locations plus timing information. These locations can be any place such as home, workplace or political party’s office, and this information can be used in an attack to breach the user’s privacy, even though names and address are anonymized in the dataset. The aim of an attack can be identifying a user’s house, workplace or behavior, and could cause a serious privacy breach of any user, where it could be possible to reconstruct social networks, knowledge of favorite visited places, political and religious views.

Collateral side effects on privacy can also be caused by initial unintentional action. For example, the geographical location, lifestyle, and other private or sensitive information may be captured through surveillance cameras installed in automated vehicles and those which initially intended to monitor criminal behaviors throughout the cities [4][11]. These vehicles can be used as surveillance tools because they are one of many smart devices that can store highly sensitive data through video and audio. Although the ownership of an automated vehicle reduces the possibility of being monitored, when used as a service it allow private and public agencies to have access to data/information. Thus, and in an extreme scenario these vehicles can be used as a platform for surveillance through the use of location tracking and audio/visual recording of passengers [11].

Another threat regarding big data is the possibility of information leaks. Icasiano and Taeihagh [12] affirms that ride-sourcing leak involving Uber data in 2016 might include trip and location data, and other data that would make possible to identify account holders and their activities. Also, because of its data-intensive nature, it is not possible for individuals to verify if their data are being used under the terms of use that they agreed, then if not leaked it still can be used in an unauthorized way.

Regarding security challenges related to data in smart cities, another potential threat is called ransomware, a malware that encrypts files to deny data access until the owner pays a ransom (typically in cryptocurrency). In a smart city context, a possible class of victims are enterprises that rely on data for operations. In a scenario where a ransomware attack targets a public transportation system, it could disable the payment infrastructure through which purchases are made, leading to financial loses or it could target the control system networks and infrastructure that actually controls the trains or manage the city's buses [32]. Besides that, intruders can leverage the structure of smart cities to create and deploy self-propagating malware, which can be disseminated across multiple connected networks [4].

V. Conclusion

This paper shed light on the discussion about data privacy on the mobility of smart cities, once the smartification process will increase availability of user-generated and user-centered information regarding urban mobility. A review was conducted and a bibliometric analysis performed. Results show that research involving data privacy in this context is recent and concentrated mainly in the United States and China. The keyword map showed the arising of new clusters from 2021, representing emerging technologies or applications, such as 5G and data fusion.

The type of data collected in smart mobility systems comprises key identifiers as name and ID numbers, quasi-identifiers as ZIP code and address, sensitive data as health condition, location and mobility patterns. Among the major treats at the individual level, identification and localization may bring life-threatening consequences. Even when some security techniques are applied to data, many datasets still present identification risks, being possible to reconstruct social networks, trajectories, favorite places, besides political and religious views. At the network level, ransomware and malware pose threats to operational and financial systems.

While smart city strategies hold great promise, it is important that data security and user privacy be taken seriously and assume a prominent role in the development of new technologies and regulations.

References

[1] V. Albino, U. Berardi and R. Dangelico, "Smart Cities: Definitions, Dimensions, Performance, and Initiatives", Journal of Urban Technology, vol. 22, no. 1, pp. 3-21, 2015. Available: 10.1080/10630732.2014.942092.

[2] R. Giffinger and H. Gudrun, "Smart cities ranking: an effective instrument for the positioning of the cities?", ACE: Architecture, City and Environment, vol. 4, no. 12, pp. 7-25, 2010. Available: 10.5821/ace.v4i12.2483.

[3] I. Docherty, G. Marsden and J. Anable, "The governance of smart mobility", Transportation Research Part A: Policy and Practice, vol. 115, pp. 114-125, 2018. Available: 10.1016/j.tra.2017.09.012.

[4] M. Sookhak, H. Tang, Y. He and F. Yu, "Security and Privacy of Smart Cities: A Survey, Research Issues and Challenges", IEEE Communications Surveys & Tutorials, vol. 21, no. 2, pp. 1718-1743, 2019. Available: 10.1109/comst.2018.2867288.

[5] R. Sinnott et al., "Privacy Preserving Geo-Linkage in the Big Urban Data Era", Journal of Grid Computing, vol. 14, no. 4, pp. 603-618, 2016. Available: 10.1007/s10723-016-9372-0.

[6] J. Ziegeldorf, O. Morchon and K. Wehrle, "Privacy in the Internet of Things: threats and challenges", Security and Communication Networks, vol. 7, no. 12, pp. 2728-2742, 2013. Available: 10.1002/sec.795.

[7] P. Lévy, *Cibercultura*. São Paulo: Editora 34, 2010a.

[8] P. Lévy, *As Tecnologias da Inteligência*. São Paulo: Editora 34, 2010b.

[9] P. Lévy, *O que é o Virtual?* São Paulo: Editora 34, 2011.

[10] D. Wei et al., "Dataflow Management in the Internet of Things: Sensing, Control, and Security", Tsinghua Science and Technology, vol. 26, no. 6, pp. 918-930, 2021. Available: 10.26599/tst.2021.9010029.

[11] H. Lim and A. Taeihagh, "Autonomous Vehicles for Smart and Sustainable Cities: An In-Depth Exploration of Privacy and

Cybersecurity Implications", Energies, vol. 11, no. 5, p. 1062, 2018. Available: 10.3390/en11051062.

[12] C. Icasiano and A. Taeihagh, "Governance of the Risks of Ridesharing in Southeast Asia: An In-Depth Analysis", Sustainability, vol. 13, no. 11, p. 6474, 2021. Available: 10.3390/su13116474.

[13] A. Hasan, Q. Jiang and C. Li, "An Effective Grouping Method for Privacy-Preserving Bike Sharing Data Publishing", Future Internet, vol. 9, no. 4, p. 65, 2017. Available: 10.3390/fi9040065.

[14] S. Fiore et al., "An Integrated Big and Fast Data Analytics Platform for Smart Urban Transportation Management", IEEE Access, vol. 7, pp. 117652-117677, 2019. Available: 10.1109/access.2019.2936941.

[15] K. Al-Hussaeni, B. Fung, F. Iqbal, G. Dagher and E. Park, "SafePath: Differentially-private publishing of passenger trajectories in transportation systems", Computer Networks, vol. 143, pp. 126-139, 2018. Available: 10.1016/j.comnet.2018.07.007.

[16] Westin A F. Privacy and freedom. Washington and Lee Law Review 1968; 25(1): 166.

[17] L. Fang, H. Wang, X. Cheng, L. Yang and S. Cui, "Mobile Privacy: Scalable Ensemble Matching for User Identification Attacks", IEEE Access, vol. 8, pp. 97243-97257, 2020. Available: 10.1109/access.2020.2995152.

[18] J. Fernández et al., "User consent modeling for ensuring transparency and compliance in smart cities", Personal and Ubiquitous Computing, vol. 24, no. 4, pp. 465-486, 2020. Available: 10.1007/s00779-019-01330-0.

[19] European Union, 2016. The General Data Protection Regulation (GDPR). [Online] Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[20] N. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping", Scientometrics, vol. 84, no. 2, pp. 523-538, 2009. Available: 10.1007/s11192-009-0146-3.

[21] VOSviewer for Windows, version 1.6.17. Developed by Nees Jan van Eck and Ludo Waltman at Centre for Science and Technology Studies, Leiden University, The Netherlands, 2021. Acessed: August, 21, 2021. Available: <https://www.vosviewer.com/>.

[22] Z. Khan, Z. Pervez and A. Abbasi, "Towards a secure service provisioning framework in a Smart city environment", Future Generation Computer Systems, vol. 77, pp. 112-135, 2017. Available: 10.1016/j.future.2017.06.031.

[23] Y. Zhao, S. Tarus, L. Yang, J. Sun, Y. Ge and J. Wang, "Privacy-preserving clustering for big data in cyber-physical-social systems: Survey and perspectives", Information Sciences, vol. 515, pp. 132-155, 2020. Available: 10.1016/j.ins.2019.10.019.

[24] N. Nedjah, R. Wyant and L. de Macedo Mourelle, "Efficient biometric palm-print matching on smart-cards for high security and privacy", Multimedia Tools and Applications, vol. 76, no. 21, pp. 22671-22701, 2017. Available: 10.1007/s11042-016-4271-8.

[25] B. Liu, S. Xie, H. Wang, Y. Hong, X. Ban and M. Mohammady, "VTDP: Privately Sanitizing Fine-grained Vehicle Trajectory Data with Boosted Utility", IEEE Transactions on Dependable and Secure Computing, pp. 1-1, 2021. Available: 10.1109/tdsc.2019.2960336.

[26] J. Sainio, J. Westerholm and J. Oksanen, "Generating Heat Maps of Popular Routes Online from Massive Mobile Sports Tracking Application Data in Milliseconds While Respecting Privacy", ISPRS International Journal of Geo-Information, vol. 4, no. 4, pp. 1813-1826, 2015. Available: 10.3390/ijgi4041813.

[27] S. Kawasaki, "The challenges of transportation/traffic statistics in Japan and directions for the future", IATSS Research, vol. 39, no. 1, pp. 1-8, 2015. Available: 10.1016/j.iatssr.2015.06.002.

[28] M. Yamin, Y. Alsaawy, A. B. Alkhodre and A. Abi Sen, "An Innovative Method for Preserving Privacy in Internet of Things", Sensors, vol. 19, no. 15, p. 3355, 2019. Available: 10.3390/s19153355.

[29] D. Eckhoff and C. Sommer, "Driving for Big Data? Privacy Concerns in Vehicular Networking", IEEE Security & Privacy, vol. 12, no. 1, pp. 77-79, 2014. Available: 10.1109/msp.2014.2.

[30] D. Goroff, J. Polonetsky and O. Tene, "Privacy Protective Research: Facilitating Ethically Responsible Access to Administrative Data", The ANNALS of the American Academy of Political and Social Science, vol. 675, no. 1, pp. 46-66, 2017. Available: 10.1177/0002716217742605.

[31] M. Abomhara, S. Yayilgan, L. Nweke and Z. Székely, "A comparison of primary stakeholders' views on the deployment of biometric technologies in border management: Case study of SMart mobILity at the European land borders", Technology in Society, vol. 64, p. 101484, 2021. Available: 10.1016/j.techsoc.2020.101484.

[32] H. Habibzadeh, B. Nussbaum, F. Anjomshoa, B. Kantarci and T. Soyata, "A survey on cybersecurity, data privacy, and policy issues in cyber-physical system deployments in smart cities", Sustainable Cities and Society, vol. 50, p. 101660, 2019. Available: 10.1016/j.scs.2019.101660.