Exploratory Analysis of Public Transportation Data of Curitiba, Brazil

Karla Kiyoko Hashiguchi, Bruno de Freitas Gai,
Daniel Fernando Pigatto and Keiko Veronica Ono Fonseca

July 7, 2020

Exploratory Analysis of Public Transportation Data of Curitiba, Brazil

1st Karla K. Hashiguchi
Univ. Tecnológica Federal do Paraná (UTFPR)
Curitiba, Brazil
karlahashiguchi@hotmail.com

2nd Bruno de F. Gai
Univ. Tecnológica Federal do Paraná (UTFPR)
Curitiba, Brazil
bruno@robosoft.com.br

3rd Daniel F. Pigatto
DAELN/PPGCA
Univ. Tecnológica Federal do Paraná (UTFPR)
Curitiba, Brazil
pigatto@utfpr.edu.br

4th Keiko V. O. Fonseca
DAELN/CPGEI
Univ. Tecnológica Federal do Paraná (UTFPR)
Curitiba, Brazil
keiko@utfpr.edu.br

***Abstract*—The dynamics between technologies and everyday life is increasingly volatile, pushing for greater convenience and ease of both trivial and complex processes. A consequence of the wide implementation of Internet of Things-based devices in smart city applications clearly reflects the mentioned dynamics by the huge amount of data to be handled. To extract useful information from such data, a clever approach to conduct data analysis is the interdisciplinary one, i.e., engineering and computer scientists as close partners of the problem domain experts. This paper addresses IoT data of a Public transportation System (PTS) aiming at contributing to the data analysis process for new insights about bus based PTS applications. Open Data of Curitiba PTS was explored looking for approaches to handle and analyze this kind of data, aiming to find important information for users, managers and planners of the public transport network in Curitiba.**

***Index Terms*—Curitiba, Internet of Things, Public Transportation, Urban Mobility.**

I. Introduction

Internet of Things (IoT) is an example of a set of technologies that are becoming accessible and that can bring more convenience into people's lives [1]. One possible area where IoT can play an important role is urban mobility: for example, by providing useful information on sensing service quality to help citizens choose better routes or transportation modals in big cities. Whether walking, cycling, car, public or individual transportation vehicle, the information can also be customized as input to mobility related applications for example, to find parking spots while avoiding worsening air pollution at specific urban regions, to reduce street congestion points, to help disabled people on choosing routes or special services; to increase the efficiency of public transportation systems (PTS) through the control of traffic lights, by sensing adaptations of buses and sidewalks to wheelchairs, or updating traffic information on display panels, among others [2].

In this paper, we present a case study of handling and analyzing IoT big data of a public transportation system (PTS) aiming at detecting urban mobility issues of a big city in South America (Curitiba). The analysis process and results provide insights for possible approaches to understand and handle IoT data aiming at improvements of the PTS quality of service. This paper is organized as follows: section II briefly presents relevant concepts regarding Urban Mobility in Curitiba and how the PTS works with IoT systems; section III describes the case study of urban mobility in Curitiba; section IV describes our methodology with the proposed approach and details about the algorithms; section V presents results and discussions; and, finally, section VI concludes the paper.

II. Urban Mobility and IoT

According to [3], public policies should prioritize public over private transportation, and investments in vehicle quality and non-polluting fuels. By improving the understanding of the user perception of the quality of available public transportation modals can positively impact on the city sustainability [4] [5]. Authors in [4] point out that urban mobility policies aim at presenting the guidelines for planners to develop projects in an appropriate way, prioritizing the well-being of users. For example, traffic management can be tackled diplomatically [6] to help build an efficient structure of an intelligent public transport management system that dynamically tracks the location of all vehicles and estimates the arrival time of the next bus at the terminal, specifically for a bus-based public transport [6]. On-demand information gathered by IoT devices can be made available to passengers and operators for better planning of trips ahead of time or quickly and efficiently reacting under non predicted events. IoT can be an enabler for urban mobility applications.

A. Urban Mobility in Curitiba

Curitiba is a city in Southern Region of Brazil that followed a master plan since the mid-1960s [7]. At that time, its Urban Mobility and Integrated Transport Plan had already stated the issue of making travel easier according to the movement of goods and people in the municipality [8]. Among several items that urban planning stands for, Collective Transportation, Road System and Land Use, seeks to integrate all the physical part of the city and driving growth in a regular and orderly manner. Created in 1965, IPPUC[1] (*Instituto de Pesquisa e Planejamento Urbano de Curitiba*) aims to monitor and coordinate the execution of the Master Plan of the city of Curitiba [9].

Authors acknowledge the Municipality of Curitiba, IPPUC and URBS.

[1] www.ippuc.org.br/

Curitiba has gained worldwide visibility for its Master Plan approach regarding the transportation system. The city plan was re-designed with structural axes that extend out from downtown to far neighborhoods ending on bus terminals. The land legislation enforces the concentration of growth and housing density along these axis and main roads, which in turn are the main routes of the public transportation system [10].

The successful implementation of Curitiba's master plan is closely related to the efficiency of the public transportation system based on its buses, bus terminals, tube stations, inter-neighborhood lines, express lanes dedicated for buses and the PTS integration with those of the metropolitan area [11]. The PTS operation in Curitiba is managed by a municipality owned company called URBS – *Urbanização de Curitiba*[2].

B. IoT and the Public Transportation System

Public transportation is a service available on a basis of sharing for the benefit of the general public [6]. It includes urban buses, trolleybuses, trams, passenger trains, ferries and fast transit, such as subways. The main reasons why people choose public transportation over other means of transport are their subsidized tariffs, environmentally friendly attributes and ease of access [2]. Most trains and buses operate on a scheduled timetable. However, the schedule adherence depends on several factors (atypical events, congestion, driver's behavior, for example) and there is uncertainty about the arrival time of buses. From the user perspective, the time of arrival and departure of each bus, a comprehensive list of bus stops, information on bus routes, etc., relate to the quality of the provided transportation service [6].

III. IoT AS AN ENABLER FOR BETTER URBAN MOBILITY SOLUTIONS – THE CURITIBA CASE OF STUDY

Public transportation in Curitiba has been adapted since the wide availability of new applications, e.g. to get a ride[3], choosing from rent a bike or scooter[4] or even asking for a car[5], which are easily accessible from smartphone apps. These applications are particularly seen as examples of IoT scenarios, mainly due to the use of devices, communication networks and control systems as shown in Figure 1.

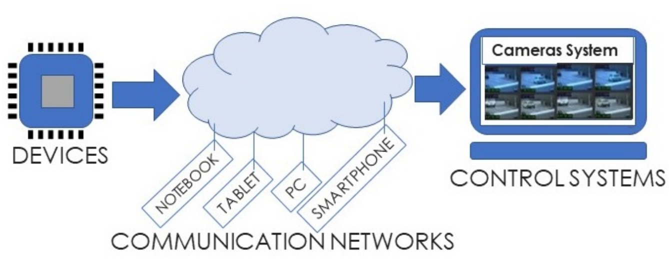


Fig. 1. Conceptual Scheme of Internet of Things.

[2]www.urbs.curitiba.pr.gov.br/
[3]www.blablacar.com/
[4]www.grin4u.ongrin.com/
[5]www.uber.com/br/pt-br/

The business model behind the aforementioned urban mobility applications [12] takes into account several factors, for example, topography, pluviometric index, choice of coverage area, public profile, population density, residential distribution, commercial activity, infrastructure of cycle tracks and paths, which are analyzed in combination. These applications make extensive use of geolocation information and collect data about the user profile, mainly for advertisement purposes. The collected information can provide hints about possible origin-destination of users [13] and guide the application provider to develop application improvements aimed at matching the user profile.

On the other hand, the inefficiency and the low quality of service of Curitiba's PTS have been pointed out as the main reasons for people to change from public to individual or shared transportation solutions[6]. It is also a fact that in Curitiba, the land use is guided by the master plan [14], which forces the PTS implementation and concentrates bus lines and people at specific points along the main PTS routes (terminals and tubes), occasionally leading to longer PTS routes (distances or travel time). Also, there is a lack of integration of the PTS with other modals that could help the user to overcome the last mile, for example from an origin (home, school etc.) to the next PTS bus stop or from the bus stop to their final destination.

Based on these hypothesis, one question arises: how the information provided by the URBS IoT devices on buses could be helpful to improve the user satisfaction to keep or attract users to the PTS? In order to answer this question we analyzed URBS available data about some particular bus stops to understand the challenge of extracting useful information from them. Our main focus is the schedule adherence of bus lines, as an important service parameter from the user perspective.

A. The PTS Information System of Curitiba

IPPUC and URBS have been providing open data about the city and the PTS operation. Applications that benefit from such information are ItiBus[7], Moovit[8] and electronic panels available in bus terminals to indicate bus departure times [15].

URBS monitors all PTS buses in real time from its Operations Center, which is only possible because all buses are equipped with on-board computers, GPS modules and other sensors [15]. Bus data includes information about the bus speed, schedule adherence, estimated trip time etc. Their tools publish relevant information such as: changes in routes or deviations from itineraries; mobile application (ItiBus) about lines, bus routes and position of stops and vehicles; taxi stops in Curitiba; individual user Transport Card information etc.

The PTS of Curitiba has 251 bus lines with 1229 buses running on a daily basis, each one sending its geolocation every minute to the URBS operation Center. All data collected by the bus is also stored on a local media and uploaded to the

[6]www.gazetadopovo.com.br/parana/transporte-aprlicativos-perda/
[7]www.urbs.curitiba.pr.gov.br/mobile/itibus5
[8]www.moovitapp.com/index/pt-br/transporte_publico-Curitiba-942

URBS database (DB) later. A copy of the daily collected data of all buses is uploaded from the URBS DB and kept at the C3SL repository as open data for research purposes. The bus geolocation data stored at C3SL has a granularity of time of approximately 10 seconds for geolocation samples, which is different than the information provided at near-real time (1 minute delay) by direct request to the URBS DB server. The higher sampling rate provides better geolocation accuracy, but also sets a bigger challenge of data handling and processing due to the volume. We opt out to work with a larger database for a better accuracy.

IV. Methodology

Data collected from the IoT system deployed to control the Curitiba's PTS operation feeds a system developed to extract useful information for the PTS users about the schedule adherence of buses from specific bus lines. The system validation was done by comparing peak hours reports from URBS.

In our work, schedule adherence is associated to the bus departure and arrival times set at the official URBS timetables[9]. These timetables are set for all bus lines associated to the time intervals showed in Table I.

Table I
PEAK TIMES.

Peak 1	From 05h00 to 08h30
Morning	From 08h31 to 11h30
Peak 2	From 11h31 to 14h00
Afternoon	From 14h01 to 16h30
Peak 3	From 16h31 to 19h30
Evening 1	From 19h31 to 22h30
Evening 2	From 22h31 to 23h59

The bus line schedule differs for workdays, weekends and special events. URBS usually checks the schedule adherence only at few bus stops (the initial, an intermediary and the final bus stop of a bus line route). We limited our analysis for workdays (Monday to Friday) that were not holidays, school vacation or with special events. Special events are those that may be joined by more than 500 people on a particular time interval and place (for example, a league soccer game at the local soccer arena). These events should be reported in advance to the local transit authority for transit arrangements prior to the reported event since they can disrupt the operation of some bus line.

Some parameters were defined to perform our analysis:
DT - Total delay time (seconds) as the sum of bus schedule delay time within the allowed time interval (positive number).
DA - Amount of delays as number of delay occurrences within the allowed time interval (positive number).
EAT - Total of early arrival time (seconds) as the sum of all bus early arrival time within the allowed time interval (negative value).
EAA - Amount of early arrivals as the number of early arrival occurrences within the allowed time interval (Negative value).
OKSch - Number of scheduled bus occurrences within the allowed time interval.
AD - Total of analyzed days as the number of analyzed days by the algorithm.
OD - Total of days with some occurrence as the number of days analyzed with some occurrence found by the algorithm.

[9]www.urbs.curitiba.pr.gov.br/horario-de-onibus

Table II
CHOSEN BUS LINES.

Bus Line ID	Name	Type	Frequency
303	Centenário-Campo Comprido	Express	each 10 minutes
507	Sítio Cercado-Pinheirinho-Guadalupe-Boqueirão	Direct	each 10 minutes
924	Santa Felicidade - Santa Cândida	Feeder	each 20 minutes

From the above listed items, we defined the following:

- Integrity Percentage (I) as the rate of total of days with some occurrence and total of analyzed days (I=OD/AD)
- Delay average (DAv) in seconds as the average value of total delay time (DAv=DT/DA)
- Early arrival average (EAAv) in seconds as the average value of early arrival time (EAAv=EAT/EAA)
- User perception (UP) as the difference of the delay and early arrival averages (UP=DAv-EAAv)

Figure 2 shows a flowchart of steps executed by our algorithm. The first step identifies which URBS data should be selected to compute the schedule adherence of a bus line. Due the number of lines and the amount of data to be handled, we limited our analysis to three bus lines that were chosen according to the main characteristic of bus lines in Curitiba (express, feeder and direct bus lines). The analysis of IoT data from the PTS of Curitiba should be able to characterize the peak time behavior of the system in terms of schedule adherence estimates. The selected bus lines and their characterization are listed at Table II. The bus line frequencies at this table are just references since frequency changes around this value for each of the time intervals of Table I.

Figure 3 shows the itinerary of bus lines 303, 507 and 924.

Hereby, bus stop refers to the place set for boarding and disembarking of users of PTS buses. Bus occurrence refers to the physical presence of the bus nearby the bus stop. Schedule adherence of Bus lines 303 and 924 was checked at the initial and final bus stops (bus terminals) and for intermediary bus stop (terminal or tube station). However, for bus line 507, the analysis of data from the intermediary bus stop revealed very challenging: in this line, the intermediary bus stops on both directions are very close from one another requiring identification of the bus direction to filter the samples before computing the schedule adherence to the specific bus stop. Due to time limitations, we do not address the analysis of intermediary bus stops that are not bus terminals, however [16] reported a solution to overcome the identification of direction also using PTS data from Curitiba.

URBS data from February to April of 2019 were collected from the C3SL site for our analysis. The data was compiled to

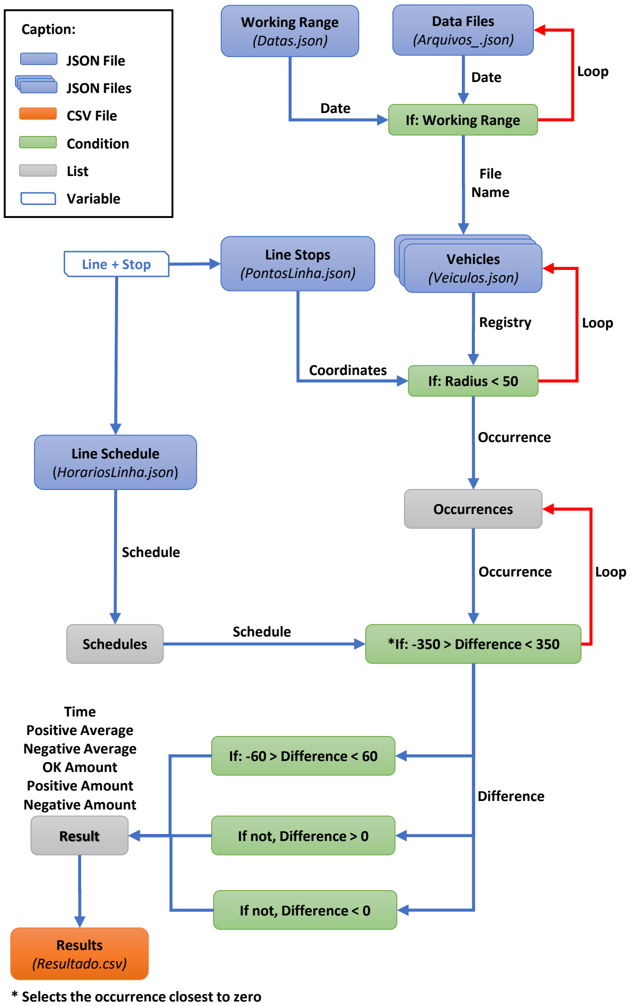


Fig. 2. Flowchart with the steps to execute our proposed algorithm.

a file (data_veiculo.json). Similarly, other files were generated for each bus line comprising the geolocation of each bus stop and the scheduled time arrival on each of them [17]. Notice that only initial and final bus stops are considered here and by the URBS official schedule. URBS metadata with JSON is a standard for PTS serial data and the files can be up to 500 MB of data, corresponding to up to 5 million of records.

Some files have their structures adapted allowing to be correctly read using Python [18]. Details on these adaptations can be found at the project repository on GitHub[10]. Consistency of records were checked for all files of the investigated period (February - April, 2019) using the algorithm available on the mentioned repository. We exemplify here one of the challenges of processing the big volume of data collected along the day for the investigated period: one desired information is the real arrival and departure time of the bus that should be compared with the scheduled time in order to compute schedule adherence. However, there are particularities regarding the geolocation information: there is a huge amount of samples repeating the same geolocation, for example, during the waiting for a green traffic light, or due to street congestion causing the bus to halt or drive very slow due. This repeated geolocation information is a redundant load that is time and processing consuming.

[10]This project's repository: github.com/BrunoGai/UrbanMobilityCuritiba

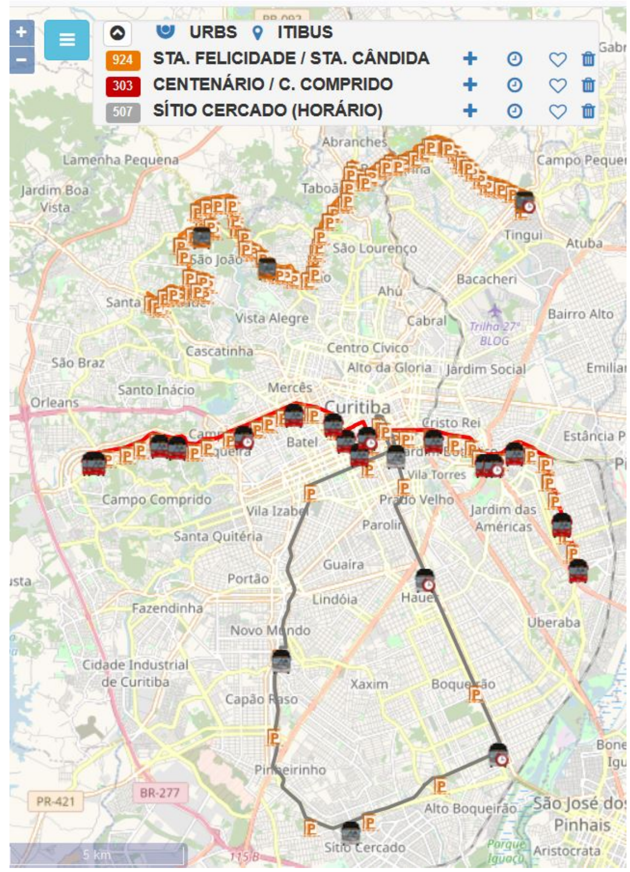


Fig. 3. Bus Line Itineraries - Source: URBS Curitiba (2020).

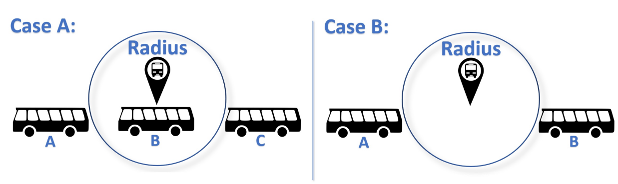


Fig. 4. Defining a Radius.

We circumvent this issue by defining a circle centered in the bus stop under investigation with a 50m radius (value chosen after several tests) represented in Figure 4 aimed to reduce the number of bus occurrences. The same figure represents in B a situation where a smaller radius was chosen and the bus occurrence at the local was missed. Larger radius leads to an increased number of redundant geolocation samples.

The schedule time and the geolocation for the investigated bus line at bus stops were done using the code listed at the repository. There is also the code used to compute the distance related to the bus occurrence and the bus stop using Haversine formula as described at [19] available on the repository.

In order to classify bus occurrences inside the aforementioned radius, a time interval of 60s was set for each sample time (set as a reference) reported inside the circle around the bus stop under investigation. Once a geolocation inside the circle is detected, all samples collected 60s before and after this moment are considered, i.e., a bus can be at most 50 meters of distance from the bus stop in some moment that cannot be more than 2 minutes apart from the sampling time. If a bus occurrence is detected inside this interval, the occurrence is classified as OK. Bus occurrences greater than 350s (11 and a half minutes) are set as delayed or early arrival depending on the official scheduled time. Figure 5 represents these conventions taking into account the frequency reported at Table II, the assumption of a possible delay could match a possible early arrival and also an assumption that the user perception of delay relates subjectively with the maximum tolerable waiting time.

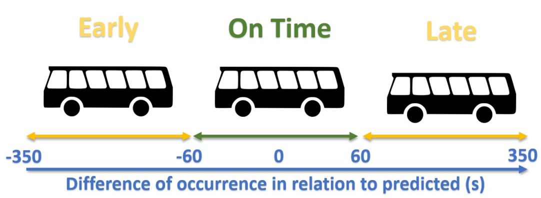


Fig. 5. Defining a Radius.

V. Results

The tables and charts presented in this section were generated by processing open data from URBS publicly available at C3SL[11]. Terminals on lines 303, 507 and 924 were selected for a closer analysis on the vehicles' performance on these routes, the user experience and what opportunities for improvement could be pointed out based on the results obtained.

Table III presents a selected range of data arranged by expected time. This is a simple example of how data is organized before being processed by our algorithm.

Figure 6 presents the variation of delays (in blue) and ahead times (in red), and Figure 7 presents the user perception for Line 924 - Terminal Santa Cândida. Similarly, Figures 8 and 9 show the same views for Line 924 - Terminal Santa Felicidade.

We computed contiguous and/or periodic incidence of delay or ahead averages greater than 3.5 minutes and selected them for a closer analysis. Figure 10 shows them in blue for several lines compared with those provided by URBS (columns of Table I. These results emphasize that not all bus lines have the same behavior regarding peak hours: they show us that bus frequency could be better distributed. From this figure, one can notice that most of the time the peak time of the bus lines are not necessarily synchronized for all the investigated bus terminals.

[11]dadosabertos.c3sl.ufpr.br/curitiba/TransporteColetivo

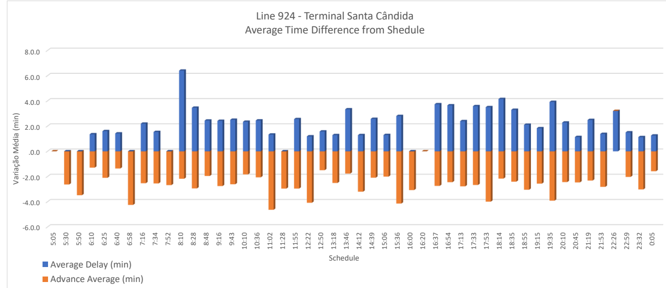


Fig. 6. Line 924 - Term. Sta. Cândida - Expected (h) vs. Average (s).

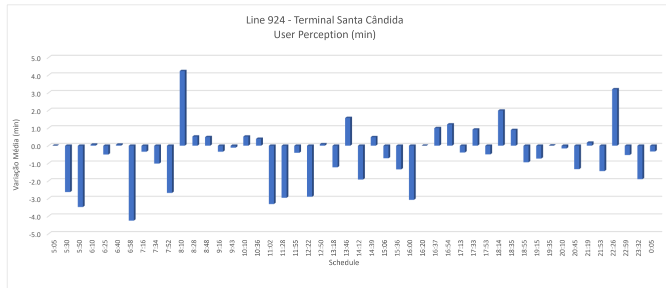


Fig. 7. Line 924 - Term. Sta. Cândida - User Perception (min).

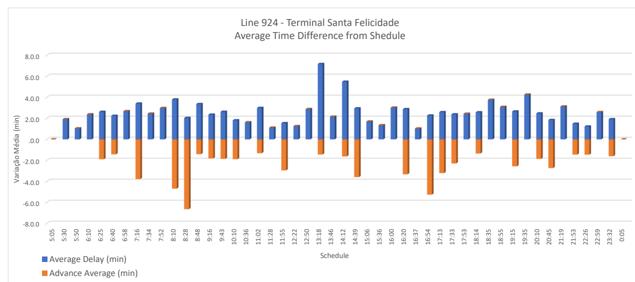


Fig. 8. Line 924 - Term. Sta. Felicidade - Expected (h) vs. Average (s).

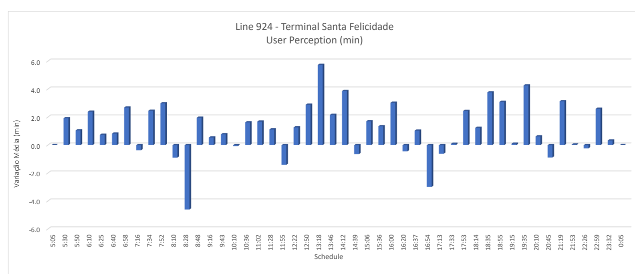


Fig. 9. Line 924 - Term. Sta. Felicidade - User Perception (min).

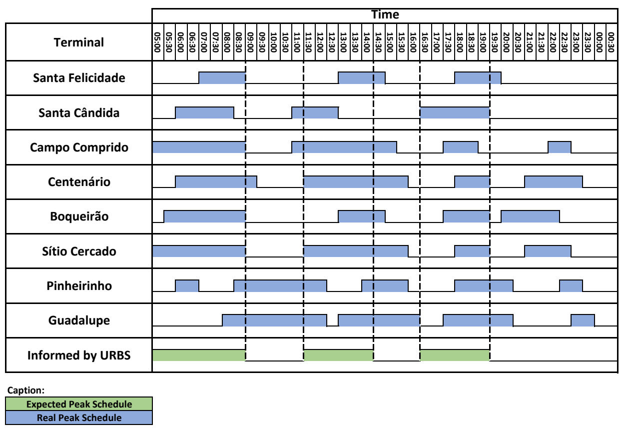
VI. Conclusions

Primarily, an exploratory, qualitative and descriptive research was done on IoT subjects applied to public transportation solutions. A case study based on available open data of Curitiba was carried out for IoT data gathered from its bus

Table III
BUS LINE ROUTE 924 - SANTA FELICIDADE.

Time	Total Delay (s)	Qnt. Delay	Total Early Arrival (s)	Qnt. Early Arrival	Qnt. OK	Total Days Found	Delay Average (min)	Early Arrival Average (min)	Total of Days Analyzed	User Perception (min)	% of Integrity
15:36	806	10	0	0	42	52	1,3	0,0	52	1,3	100%
16:00	1.997	11	0	0	36	47	3,0	0,0	52	3,0	90%
16:20	2.759	16	-398	2	34	52	2,9	-3,3	52	-0,4	100%
16:37	186	3	0	0	49	52	1,0	0,0	52	1,0	100%
16:54	2.722	20	-315	1	29	50	2,3	-5,3	52	-3,0	96%
17:13	2.176	14	-384	2	36	52	2,6	-3,2	52	-0,6	100%
17:33	856	6	-137	1	45	52	2,4	-2,3	52	0,1	100%
17:53	2.927	20	0	0	32	52	2,4	0,0	52	2,4	100%
18:14	3.086	20	-241	3	29	52	2,6	-1,3	52	1,2	100%
18:35	3.843	17	0	0	34	51	3,8	0,0	52	3,8	98%
18:55	6.115	33	0	0	19	52	3,1	0,0	52	3,1	100%
19:15	1.753	11	-460	3	38	52	2,7	-2,6	52	0,1	100%
19:35	767	3	0	0	48	51	4,3	0,0	52	4,3	98%
20:10	2.814	19	-776	7	26	52	2,5	-1,8	52	0,6	100%

Fig. 10. Peak times by terminal.



based PTS. Our study revealed challenges of processing big data and provided insights on the PTS operation.

Our work contribution towards approaches to understand the user perspective on the PTS quality of service and possible applications using quantitative and qualitative analysis. As a proof of concept, our analysis found out important information for users, operators, managers and planners of the public transport network in Curitiba. The development of a real-time passenger information system that integrates transport information with collective-vehicle locations is a potential future work based on our methodology, which might be delivered via mobile apps to the end user. Such implementation leads to improved information available to help users and bus companies by reducing the waiting times at the pick up points and delivering a more efficient service.

Future work includes the integration of a methodology for mapping the direction of a bus and, consequently, the line bound (as proposed by [16]). A further integration between the Customer Service Center and PTS could be carried out to identify potential targets for improvement.

REFERÊNCIAS

[1] F. Lacerda and M. Lima-Marques, "Da necessidade de princípios de Arquitetura da Informação para a Internet das Coisas," *Perspectivas em Ciência da Informação*, vol. 20, no. 2, pp. 158–171, jun 2015.

[2] A. Melis, M. Prandini, L. Sartori, and F. Callegati, "Public transportation, iot, trust and urban habits," in *Internet Science*, F. Bagnoli, A. Satsiou, I. Stavrakakis, P. Nesi, G. Pacini, Y. Welp, T. Tiropanis, and D. DiFranzo, Eds. Springer Intern. Pub., 2016, pp. 318–325.

[3] J. Wilheim, "Mobilidade urbana: um desafio paulistano," *Estudos Avançados*, vol. 27, no. 79, pp. 7–26, 2013.

[4] C. F. Bachendorf, "Inteligência, sustentabilidade e inovação nas cidades: uma análise da mobilidade urbana de Pato Branco - PR. 2018," *Universidade Tecnológica Federal do Paraná*, p. 152, 2018.

[5] J. Gehl, *Cidades para pessoas*, 3rd ed., Perspectiva, Ed., S.Paulo, 2015.

[6] S. H. Sutar, R. Koul, and R. Suryavanshi, "Integration of Smart Phone and IOT for development of smart public transportation system," in *2016 International Conference on Internet of Things and Applications (IOTA)*. IEEE, jan 2016, pp. 73–78.

[7] Curitiba, "Plano diretor de Curitiba," 2015. [Online]. Available: http://www.ippuc.org.br/mostrarpagina.php?pagina=310

[8] Silva, "Proposta de benchmark para simulações de roteamento de dados em redes veiculares ad hoc," 2015.

[9] IPPUC, "Plano Diretor de Curitiba é da Década de 1960," Instituto de Pesquisa e Planejamento Urbano de Curitiba, 2014. [Online]. Available: http://www.curitiba.pr.gov.br/noticias/primeiro-plano-diretor-de-curitiba-e-da-decada-de-1960/32276

[10] J. Rabinovitch, "Innovative land use and public transport policy: The case of curitiba, brazil," *Land Use Policy*, vol. 13, no. 1, pp. 51–67, 1996.

[11] A. Cinquina and B. Holmberg, "Sustainable public urban transport systems: The case of curitiba," *Lunds Universitet*, 2006.

[12] Camila Machado, ""Corajoso foi vir para a América Latina", diz diretora da patinetes Grin," 2019. [Online]. Available: https://www.gazetadopovo.com.br/haus/inovacao/paula-nader-grin-yellow-explica-que-servico-nao-e-inovador/

[13] T. Braz, M. Maciel, D. G. Mestre, N. Andrade, C. E. Pires, A. R. Queiroz, and V. B. Santos, "Estimating inefficiency in bus trip choices from a user perspective with schedule, positioning, and ticketing data," *IEEE Trans. on Intelligent Transportation Systems*, vol. 19, no. 11, pp. 3630–3641, 2018.

[14] E. da Silva, M. de Oliveira Rosa, K. Fonseca, R. Luders, and N. Kozievitch, "Combining k-means method and complex network analysis to evaluate city mobility," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2016, pp. 1666–1671.

[15] Urbanização de Curitiba (URBS), "URBS - Urbanização de Curitiba," 2019. [Online]. Available: https://www.urbs.curitiba.pr.gov.br/

[16] E. Manika, J. Alves Jr., E. Wille, K. Fonseca, and A. Vendramin, "Um esquema automatizado de mapeamento de mapas com importação de dados do transporte público para o sumo," in *Anais do Simpósio Bras. de Telecomunicações e Processamento de Sinais*. SBRT, 2019.

[17] URBS, "URBS, Urbanização de Curitiba." [Online]. Available: https://www.urbs.curitiba.pr.gov.br

[18] A. A. E. Abdel and A. Kannan, "JSON encryption," in *2014 Intern. Conf. on Computer Communic. and Informatics*. IEEE, 2014, pp. 1–6.

[19] S. Hartanto, M. Furgan, A. Siahaan, U. Putera, and W. Fitriani, "Haversine method in looking for the nearest masjid," *International Journal of Engineering Research*, vol. 3, pp. 187–195, 08 2017.