Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Public Transport Occupancy Estimation using WLAN Probing and Mathematical Modeling

Túlio Vieira[a], Paulo Almeida[a], Magali Meireles[b], Renato Ribeiro[c]

[a]*Intelligent Systems Laboratory, CEFET-MG, Av. Amazonas, 7675 – 30510-000 – BH, MG, BRAZIL*
[b]*Institute of Mathematical Sciences and Informatics, PUC-Minas R. Walter Ianni, 255 – 31.980-110 – BH, MG, Brazil*
[c]*Transport Engineering Department, CEFET-MG, Av. Amazonas, 7675 – 30510-000 – BH, MG, BRAZIL*

Abstract

The rapid urbanization and consequent metropolization of cities are phenomena present in many countries around the World. This increment in urban population causes changes in personal relationships and in the physical structure of cities. Among the transformations caused by metropolization, the most noteworthy are the changes in urban mobility systems. In emerging countries, the problems generated by this situation are exacerbated, since the social, environmental and, mainly, economic characteristics make the solution of problems related to urban transportation very difficult. As consequences, it is common to have daily commutes happening mostly by road transport, absence of effective public policies and poor transport data acquisition. Common ways to achieve quality in public transportation systems are to estimate the public transport occupancy (PTO) rate and the origin-destination matrix for a given area. This paper presents a formulation and study for the use of a low cost and reduced computational complexity system, capable of counting in real time the amount of passengers inside a public transportation unit (bus or metro composition). This system performs passenger counting through users' smartphones electronic fingerprints and uses a mathematical model to adjust the acquired raw values. The applicability of this approach is demonstrated through case studies carried out in 3 different bus lines in the city of Belo Horizonte, Brazil, as an alternative to collect field data for the management of public transportation. It is demonstrated that the suggested system presents good accuracy and is a reduced complexity alternative for the estimation of PTO in emerging countries.

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY.

Keywords: Public Transport Occupancy; Intelligent Transport Systems; WLAN Probing; Mathematical Modeling

1. Introduction

The rapid growth of cities in recent years, especially in emerging countries, has been accompanied by a phenomenon known as metropolization. According to Méo (2008), metropolization occurs due to the concentration of people in urban areas, causing cities to increase their absolute population considerably. However, this growth of cities, especially in developing countries, is not accompanied by any improvement in urban infrastructure that meets the needs of all citizens of the city. Campolina and Vieira (2016) report 4 different issues faced by emerging countries after the phenomenon of metropolization. Among these issues, Campolina and Vieira (2016) highlight the

Peer-review under responsibility of WORLD CONFERENCE ON TRANSPORT RESEARCH SOCIETY

problems generated by the poor public transport system. The poor quality transport system generates brutal urban traffic congestion, increases commuting and travel times, increases accident rates, resulting in high death rates. According to DATASUS official calculations, in 2014, in a rank of external causes of death, the number of deaths caused by traffic accidents took second place. As it can be seen, public transport plays a key role in the lives of city dwellers.

Another common feature in developing countries is the increase in the number of motorcycles. This significant increase contributes to the reduction of the quality of collective transportation and generates great inconveniences for urban life. In Brazil, for example, between 2004 and 2017, the fleet of motorcycles grew 270% and already represents 27% of the fleet of motor vehicles (Sindipeças, 2018). According to DPVAT (2016), accidents involving motorcycles accounted for 76% of the compensation paid to victims of traffic accidents in the year 2016. In addition, motorcycles are the means of transportation most involved in all types of traffic accidents. Thus, these data show that modifying the current model of collective transportation is essential to attract more users, reduce personal transportation, and improve the quality of life in Brazilian cities.

According to a research carried out by the national association of public transport enterprises (NTU, in the Portuguese abbreviation), in several capitals of Brazil, in 2017, the number of passengers using public transport has been declining in recent years. These data are at odds with what happens in more developed countries. This fact can be explained by the poor quality of the service provided, the value of the paid tickets and the increase in the average time spent to reach the destination when compared to the time spent by private transportation. One consequence of this increase in travel times is the increase in the number of cars on the streets, which further increases travel time.

An Origin-Destination Survey (ODS) characterizes urban mobility and makes it possible to analyze changes in patterns of commutes and travels for a given metropolitan area. For the Metropolitan Region of Belo Horizonte, the most recent ODS were realized in the years 2002 and 2012. Figure 1, based on these surveys and other indicators, shows that, in this period, automobile and motorcycle fleets grew up enormously (112% and 278%, respectively, representing an increase of 92.7% in the motorization rate). During the same period, the number of trips per inhabitant grew up by 57%, a growth mainly driven by motor and motorcycle trips, which increased by 174% (Ribeiro, 2015). Taken together, these data demonstrate a reversion in the use of the main modes of transport, with a continuous increase in the use of individual transport, to the detriment of public, collective transportation.

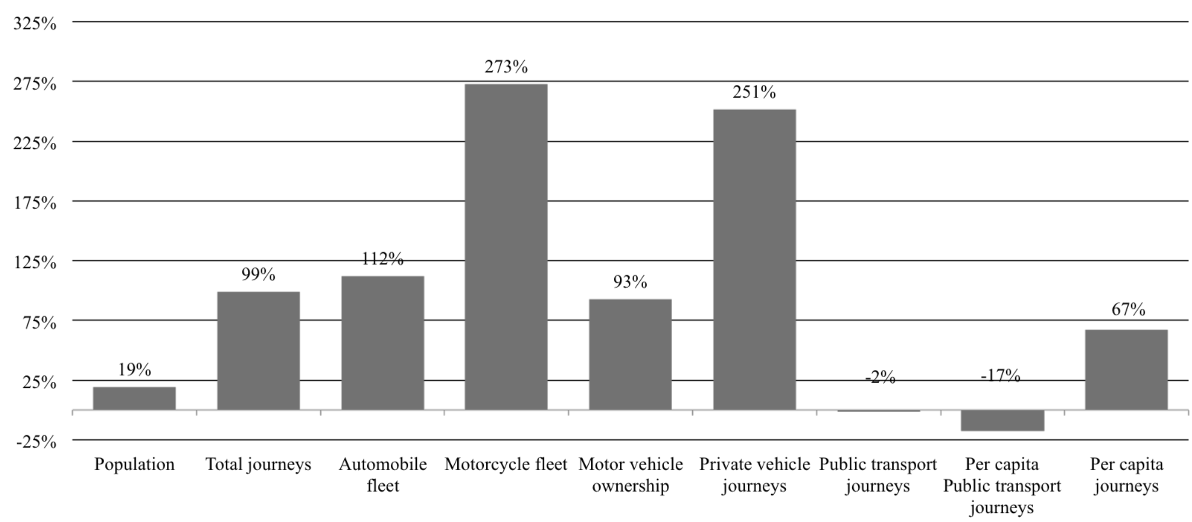

Fig 1: Population growth and other mobility indicators in Belo Horizonte, Brazil, between 2002 and 2012.

The main consequence of this reversal tendency is an expenses increase with urban commutes, an increase of road congestions and, consequently, growth in average population travel times. With bigger travel times, more passengers are avoiding using public transport, which directly impacts the life quality in such cities. Figure 2 shows that, in 2002, average travel time in public transportation was 1.88 times greater than private vehicle travel, and that, in 2012, this number increased to 2.38 in the city of Belo Horizonte.

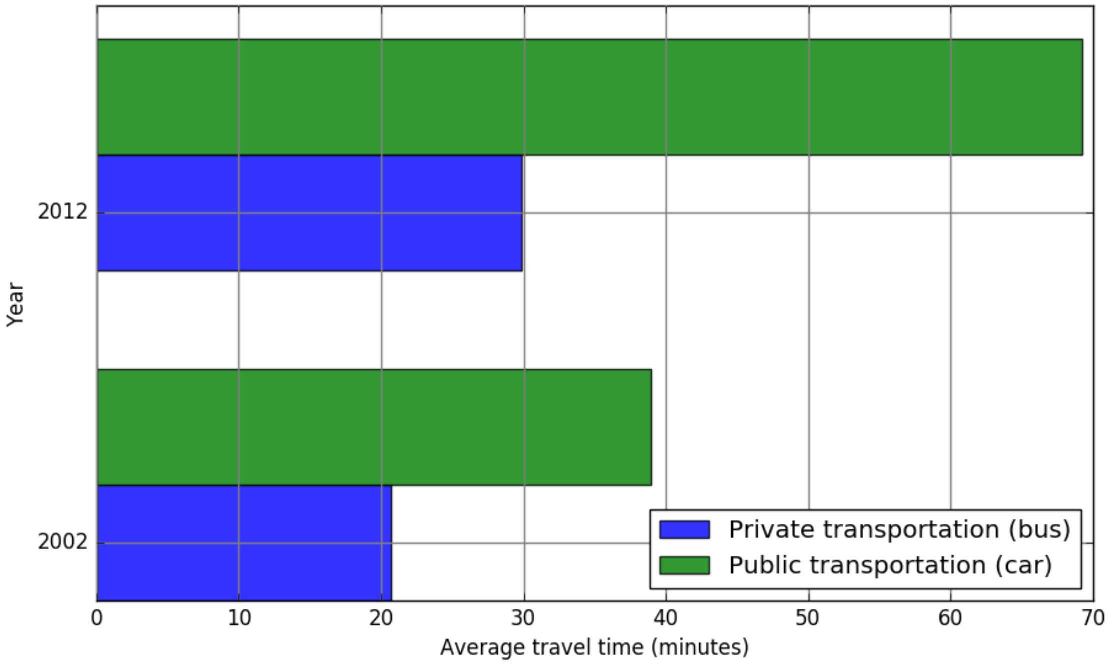

Fig 2: Comparison between travel times of public and private transport in Belo Horizonte, between 2002 and 2012.

Public policies aimed at improving the quality of public transport are very important to make this mode of transport more attractive, especially in emerging countries. One of the major problems facing emerging countries is the lack of reliable historical data that help build models and better understand the complex urban movement system. Litman (2016) indicates that quality public transportation can increase physical fitness, mental health and reduce health care costs.

Collecting and analyzing data is critical to provide a good quality public transportation system. One of the main problems faced by public transport managers in emerging countries is the difficulty in obtaining a reliable database that could assist in decision making. Whether to introduce new services, to analyze quality of existing systems or to compare different strategies that can improve the quality of public transport, database represents a valuable key to make up Intelligent Transportation Systems (ITS). Thus, due to the low reliability of historical data, developing real-time data collection systems can be an interesting strategy to help improving the quality of existing information.

One of the essential tools that can be used to achieve high quality in the public transportation system is to estimate the public transport occupancy (PTO) rate. PTO estimates the number of people on a bus, segmented along a bus line. Usually, it is obtained manually, which requires high human resource and investment demand. Accurate PTO information is used to get passenger flows, to create origin-destination matrices, to reduce costs of bus passenger transportation and to plan the transportation based on peak occupancy rate (POR). According to Heidtman et al. (1997), as POR is the best indication of the suitability of a provided service, it is extremely necessary to improve the quality of the public transportation.

In addition to manual methods, which are costly and inaccurate, there are several ways to obtain PTO rates. For instance, Loga et al. (2016) use Weight-In-Motion (WIM) system for estimating public transport vehicles occupancy rate. The results show that this approach can be applied to find PTO. However, employing WIM to find PTO requires a large financial investment. Another way to obtain PTO rates is through Wi-Fi Media Access Control (MAC) address scanning and processing.

Introduced in 1999, Wi-Fi is nowadays the most popular type of wireless networking technology used to connect to the Internet. Wi-Fi was implemented by the IEEE 802.11 project. There are a wide variety of devices that use this technology, such as smartphones, tablets and laptops. With the development of Internet of Things (IoT), many other devices and home appliances (fridges, air conditioning, cooker etc) are embedding Wi-Fi to connect to the Internet.

Smartphones or other mobile embedded Wi-Fi devices always try to connect to an Access Point (AP). This happens in two ways: by probing for any known AP, which has been connected previously, or by broadcasting a probe request to find any available AP to connect to. This probe request contains a specific identifier, which is known as MAC address. A probe request can be captured in a controlled environment and the device MAC address, that is globally unique, can be used to identify a specific person.

There are many systems that capture Wi-Fi probe requests as a way of identifying movements in ITS. One of the most effective applications is presented by London (2018). There, MAC addresses and other sources of data such as closed-circuit television cameras (CCTV) and sensors are used to automate passengers counting on buses. However, to count the number of passenger by using Wi-Fi in that approach, the passengers need to connect to an AP provided by Transport For London (TfL). In the approach here proposed, the passengers flow within a bus is counted by MAC probe requests, without providing any AP to receive Wi-Fi connections. Another system that uses IoT devices to constantly collect data on pedestrian and vehicle traffic is iSensing. iSensing is an intelligent mobility analysis platform that provides tools to improve transportation flows. In London, a system called iBus was implemented using this technology. This system uses some IoT technologies like Radio Frequency Identification (RFID) to provide unique bus stop identifiers and to record arrival and departure times at each bus stop. The collected data are used to map the transportation flow in London (Hardy, 2009). Mikkelsen et al. (2016) show that Wi-Fi probe requests, used to estimate public transportation occupancy, represent a feasible approach. Despite of the results achieved, Mikkelsen et al. (2016) used an algorithm with low complexity and only RSSI was applied. Here, we suggest a different algorithm for capturing and handling the captured MAC addresses. In our approach, we can avoid two main problems listed by Mikkelsen et al. (2016), as we will discuss later. The approach suggested here manages to estimate PTO rates with low cost and high accuracy, based only on the location of the bus stop and the MAC address scanning.

The main objective of this work is to present a low cost, high reliability and low complexity real time passenger counting system to estimate PTO with high accuracy. Thus, it is expected to be able to improve quality of public transport in emerging countries, since high public transport investments are not seen as a priority on these countries. Thus, developing low-cost tools that help providing quality public transport is critical to improve the quality of life in cities of emerging countries.

The remaining of this paper is organized as follows: Section 2 presents some background and the main model proposed to estimate PTO, Section 3 proposes a case study in Belo Horizonte, Brazil, and shows up practical experiments to validate the proposed model, discussing obtained results, and Section 4 concludes the report, summarizing this approach and suggesting ways to improve the overall accuracy already achieved.

2. A Low Cost Approach to Estimate PTO

Ojeda et al. (2013) divide traffic flow forecast into three groups: 1) approaches that use just historical data; 2) strategy that uses current information and 3) approaches that use both, historical and current information to traffic flow forecast. Because historical data is not available, we developed a mathematical model in conjunction with the current MAC data provided by the developed system and applied the Kalman Filter to fit our suggested model.

The Kalman Filter is an iterative mathematical process that uses mathematical modeling and sensor data to quickly estimate the real value of a variable. The Kalman Filter is widely used in engineering because it is, in some cases, indispensable for the control of some dynamic processes. Because it is an iterative process, the Kalman Filter has characteristics that aid in the computational implementation, which facilitates its use in automatic systems. Some classic applications of the Kalman Filter are the monitoring and control of complex dynamic systems, such as continuous manufacturing processes, aviation, navigation and spacecraft (Grewal et al., 2007).

The process of estimating the real value of a variable, performed by the Kalman Filter, occurs in two distinct and complementary steps. The first step, represented by equations 1 and 2, is known as the Prediction Step. At this stage, the mathematical model that describes the behavior of the system is used to estimate its next probable state, see Eq. 1. After, Eq. 2 can be seen as the representation of variance in the process of estimating the future state by means of the created mathematical model.

$$x_k = \mathbf{A}x_{k-1} + \mathbf{B}u_k + w_k \tag{1}$$

$$P_k = \mathbf{A}P_{k-1}\mathbf{A}^T + Q_k \tag{2}$$

After estimating the state in time (k), the step known as Update Stage begins. In this step, the estimated state in the prediction step is adjusted through the measurement obtained by the sensors. Since, in a measurement process, there is inherent error (noise) in the sensors, this adjustment occurs through Eq. 3, which represents the Kalman Gain. It is by means of the Eq. 4 that the prediction of the state is adjusted. As it can be seen in Eq. 3, based on the relation between the errors attributed to the prediction process (see Eq. 2) and the noise in the measurements, the future state value of the system is adjusted. In Eq. 5, index "km" represents the measurement at time "k". Through this iterative process, a Kalman Filter can quickly estimate the future state of a dynamic system. Fig 3 shows an overview of how a Kalman Filter works.

$$K_k = \frac{P_k\mathbf{H}^T}{\mathbf{H}P_k\mathbf{H}^T + R} \tag{3}$$

$$x_k = x_k + K_k[y_k - \mathbf{H}x_k] \tag{4}$$

$$y_k = \mathbf{C}y_{km} + \mathbf{Z}_k \tag{5}$$

$$P_k = (\mathbf{I} - K_k\mathbf{H})P_k \tag{6}$$

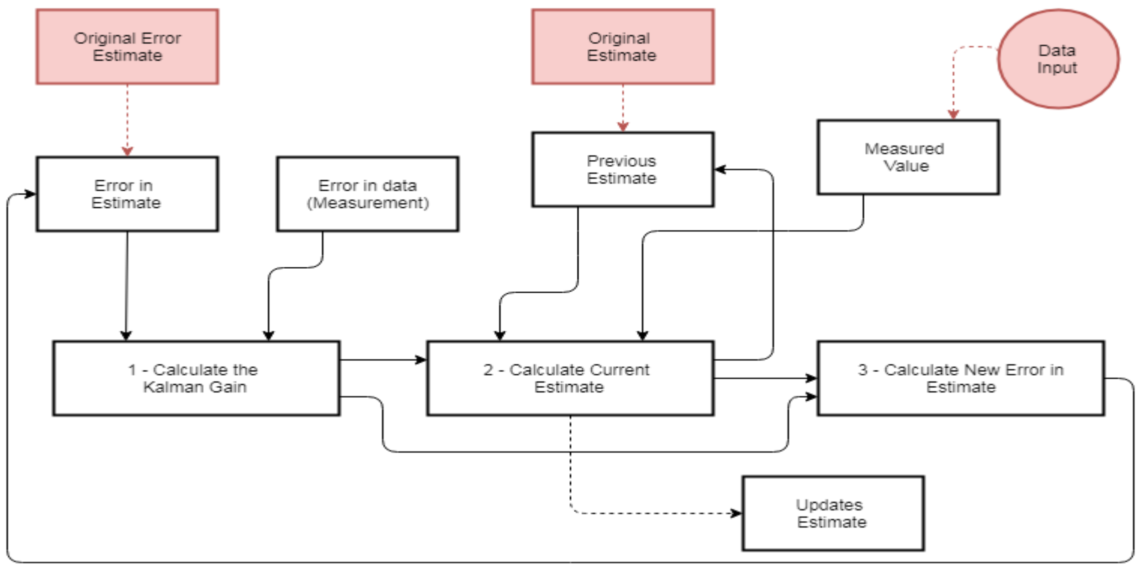

Fig 3: Overview of a Kalman Filter flowchart.

The approach introduced in this work to measure the flow of passengers through a bus is divided into two different stages, as can be seen in Fig 4. In this figure, the sensor and server stages can be clearly identified. At the sensor stage, two different kinds of data are collected and files A and B are created to store them. These files contain MAC addresses, extracted from the captured probe requests, and the location data captured from GPS system, respectively. These files are used as server stage inputs. In the server stage, by using the location database of each of the bus stops and the real time data of the bus route, obtained via GPS, it is possible to estimate the moment a bus arrives and leaves passenger stops. Once a bus is inside and outside of a bus stop, timestamps are compared to timestamps of each captured MAC and thus passenger flow at each bus stop can be calculated.

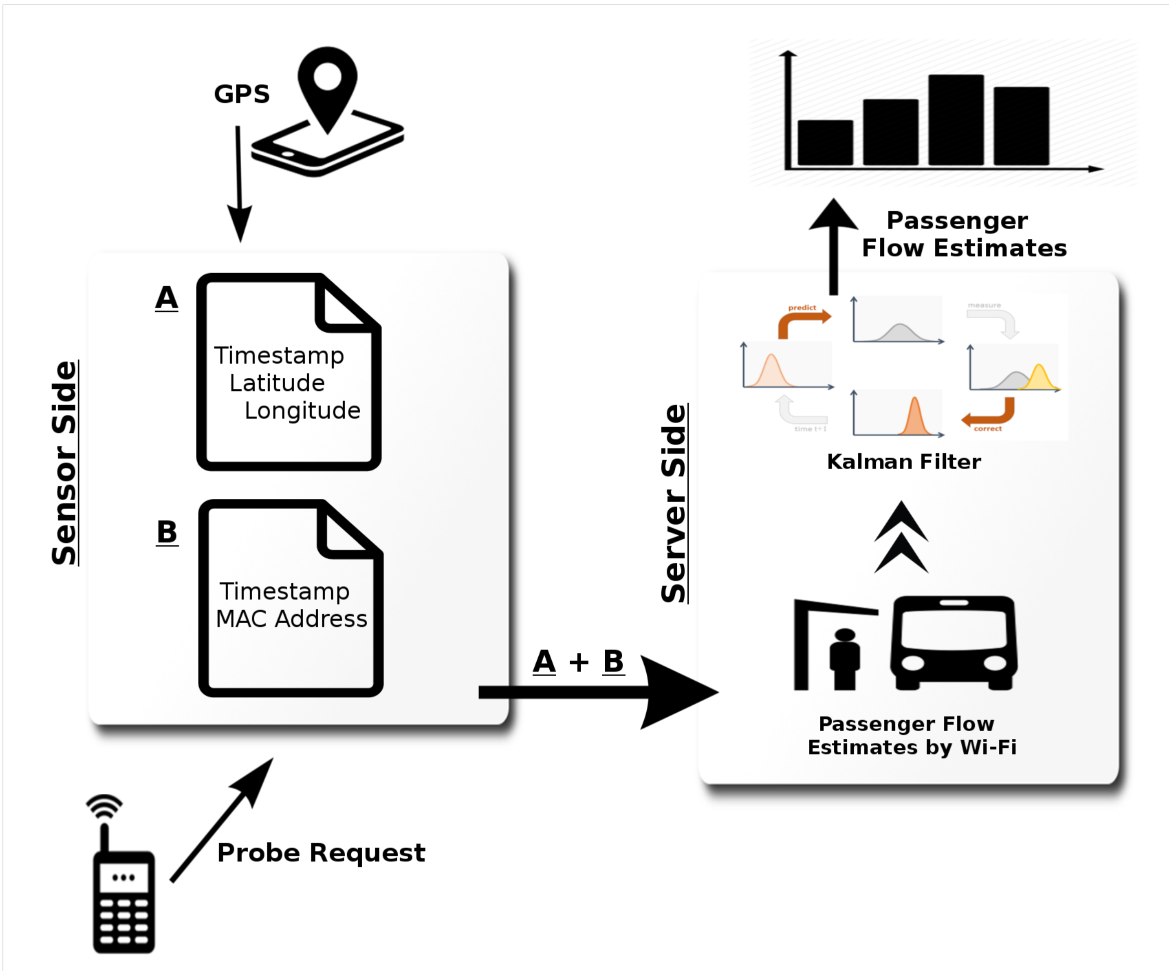

Fig 4: Main scheme for the proposed model.

It is obvious that, to capture MAC addresses, there is a need for enabled Wi-Fi. Besides public transport users not using a smartphone, not all people leave their smartphone's Wi-Fi enabled while traveling, and therefore an inherent error in calculating the passenger flow using just this strategy is expected. In our approach, a mathematical framework is used to minimize this measurement error. In order to deal with this, an autoregressive model and a Kalman Filter are used. The autoregressive model used to represent the state transition is presented in Eq. 7.

$$x_k = \alpha x_{k-1} + norm(2,3) \tag{7}$$

To fit the model to collected data, “α” is determined by the difference between the data obtained by MAC measurements and the actual measured flow values. This difference can vary for each bus line, so it must be estimated for each line, prior to extensive data acquisition. However, for the case studies performed in our research so far, a fixed value was adopted for “α”, in order to verify the robustness of the model. The portion of Eq. 7, which updates the autoregressive model, uses noise with a normal distribution with mean 2 and standard deviation 3. These values for mean and standard deviation were chosen after the collected flow data. For a better tuning of the model, historical flow data must be used. However, a reliable database was not available in the beginning, so a few collections of passenger flow values were performed to estimate “α”, mean and standard deviation values for the first practical experiments conducted. Examples of the impacts of choosing “α” values will be discussed in the next section.

3. Case study and Practical Experiments

In order to verify the applicability of the proposed approach, three bus lines were selected from the city of Belo Horizonte, capital of Minas Gerais State, Brazil. PTO results obtained using our approach, outlined in Fig. 4, are compared to the passenger flow information obtained by manual counting during trips on each of the three chosen bus lines. All these three bus lines pass through the central region of Belo Horizonte. In the central region of Belo Horizonte, the largest number of commercial and services points in the metropolitan-area is concentrated. Therefore, the passenger flow to this region is very high.

Fig. 5 shows an example of the path traveled by one of the lines studied. The central area is represented by the yellow cycle. The purple dots are the bus stops found by the algorithm running in server-side. These points were found using the Haversine distance, which is used to estimate the stopping points on the bus route. The real bus stops were found by using the database.

Fig 5: An example of a bus route in Belo Horizonte, Brazil.

Fig. 6 shows PTO estimated using only Wi-Fi MAC address scanning, without Kalman Filter adjustment. As can be seen in Fig. 6a, the proposed model, using only Wi-Fi MAC, always estimates a number of passengers during the whole trip smaller than actual number. The number of disabled Wi-Fi smartphones in buses can explain this fact, besides number of passengers who do not use a smartphone at all. In Fig. 6b, it can be seen that the proposed model, in some bus stops, estimates a grater number of passenger than the actual one.

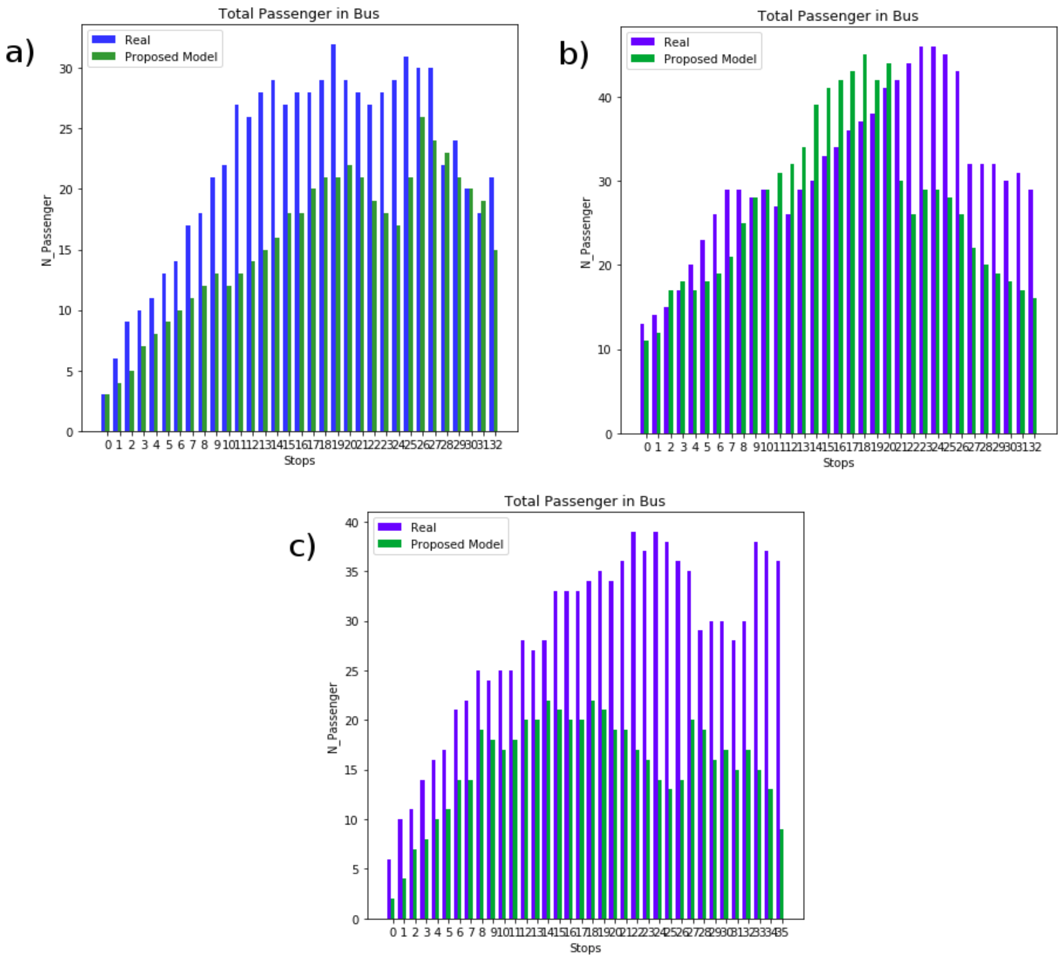

Fig 6: Bus load Estimation: (a) case study I; (b) Case study II; (c) Case study III.

In addition, we can note, in Fig.6b, that if we shift the estimated flow of passengers by the proposed model to the right, we could reduce the difference between actual and estimated PTO. This happened because, for this case study, the actual measured PTO did not use the location of the bus stops, so we could not match the same bus stops obtained by our proposed model with the actual PTO. Thus, we used the data sequence by the actual PTO to compare with the proposed approach.

As it can be seen in Fig. 6, there is a significant difference between the data obtained by only counting the MACs in relation to the actual values (obtained by manual counting). As expected, there is the possibility that the public transport user may not always be Wi-Fi enabled. Thus, to improve the accuracy of the proposed model, it is necessary to perform a correction in the counts of MAC addresses.

In order to adjust the automatic passenger flow count, a Kalman Filter is applied. Data collected by the proposed system are used to feed an autoregressive model, as presented in Eq. 7. Afterwards, a Kalman Filter is applied to adjust PTO estimates given by the model. In the three case studies, presented in Fig. 7, there is a common pattern of behavior defining the variation in the number of passengers during a trip. At the beginning of a journey, there is a growth in the bus load, until it reaches a maximum value, from which the load begins to decrease. This maximum value occurs in the places prior to the bus entering the central region of Belo Horizonte. As previously stated, the central region has the highest concentration of commercial and services points in the city. Thus, most users of public transport, for the studied lines, have as final destination the central area of Belo Horizonte.

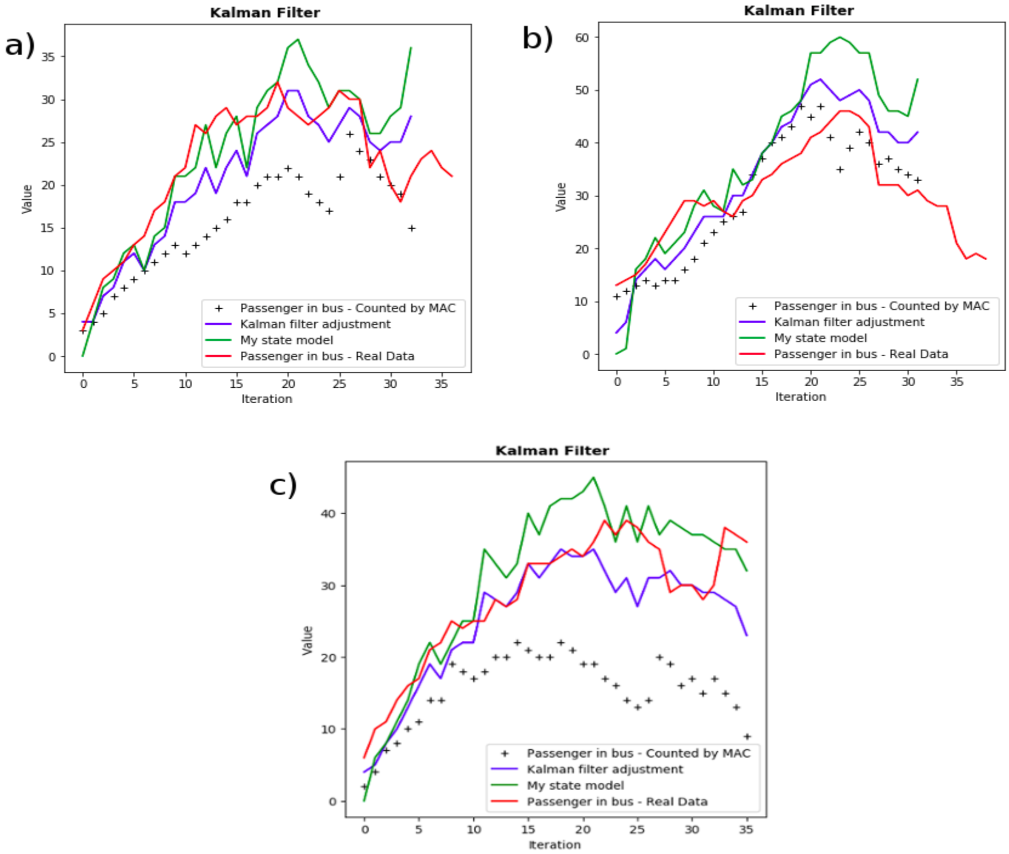

Fig 7: Kalman Filter adjustment: (a) case study I; (b) case study II; (c) case study III.

Fig. 7 shows that the adoption of the Kalman Filter for adjustment of PTO estimates guarantees a clear reduction in the difference between values estimated by MAC address count and values found by manual count of passenger flow. This important improvement occurs due to the adoption of the state model present in Eq. 7 and the corrections generated by the Kalman Filter. Since the model of Eq. 7 embodies a portion of error generated by Wi-Fi unavailability, the model presents relevant variation for the forecast of passengers' flow in each of the bus stops. This portion is implicitly used by Kalman Filter to introduce statistical uncertainties to the overall framework for forecasting the flow of passengers and thus to portray with more fidelity the characteristics of variation of PTO during a trip.

With the adoption of these two steps (sensor stage and server stage), the proposed approach was able to estimate the flow of passengers during a bus route with good accuracy, at a low cost and using a low complexity hardware tool. As reported by Mikkelsen et al. (2016), it is demonstrated that the strategy of measuring the flow of passengers through MAC can be used as a useful tool for the management of public transport. In addition, the software strategy proposed in this work is able to treat the sensitivity of the sensor data to the values of the parameters of the autoregressive model, and to increase the accuracy of final results without using expensive hardware or significantly changing the computational complexity of the software algorithm, which are two of the challenges evidenced by the strategy described by Mikkelsen et al. (2016).

4. Conclusion

One of the biggest problems faced by emerging countries that have undergone the process of metropolization is being able to deal with problems generated by the considerable increase of urban population. Poor quality of transport services compromises the balance of urban life and affects the health of inhabitants. In order to assist in the planning of the urban transport system, it is necessary to carry out research to assist in evaluating the transportation service provided, facilitating the management and introduction of new transport models. Thus, low-cost approaches that use data collected in real time may be an alternative to assist urban transport management.

The proposed low-cost automatic data collection system utilizes smartphone MAC address data and a mathematical framework including Kalman Filter adjustment to estimate PTO. The results show that the suggested approach can be used as an alternative tool to the traditional manual surveys done to estimate PTO. Improvements due to Kalman filtering dramatically reduced raw passenger flow prediction errors, around 300%. Using this framework, bus load estimation can be done with high accuracy, low cost and low complexity. This suggests that the proposed system can be used extensively to estimate PTO, for big metropolitan areas.

The public road transport service has, as central characteristics, implantation low cost and high flexibility. The existence of an approach such as the one here proposed strengthens this characteristic and contributes to the increase of service quality offered and to the creation of new models of services, including, e.g., transport on demand.

In order to improve the accuracy of the proposed approach, mathematical model parameter tuning through historical data can be realized. Thus, through computational intelligence algorithms, these parameters can be better estimated and the quality of the model can be further increased.

Acknowledgements

The authors would like to thank CEFET-MG for the infrastructure used by this project and also CAPES, CNPq and FAPEMIG for the financial support.

References

Calabrese, Francesco, Giusy Di Lorenzo, Liang Liu, e Carlo Ratti. “Estimating Origin-Destination flows using opportunistically collected mobile phone location data from one million users in Boston Metropolitan Area.” IEEE Pervasive Computing 10 (2011): 36-44.

Campolina Diniz, Clélio, e Danilo Jorge Vieira. “Brazil: Accelerated metropolization and urban crisis.” Area Development and Policy (Taylor & Francis) 1 (2016): 155-177.

Grewal, Mohinder S., Lawrence R. Weill, e Angus P. Andrews. Global positioning systems, inertial navigation, and integration. John Wiley & Sons, 2007.

DPVAT. Seguradora Líder. O Seguro do Trânsito. Boletim estatístico year 06, v.4, jan/dez. 2016.

Hardy, Nigel. “iBUS Benefits Realisation Workstream: method & progress to date.” 16th ITS World Congress and Exhibition on Intelligent Transport Systems and ServicesITS AmericaERTICOITS Japan. 2009.

Heidtman, K., B. Skarpness, e C. Tornow. “Improved vehicle occupancy data collection methods.” Tech. rep., Office of Highway Information Management, Federal Highway Administration, 1997.

Litman, Todd. Evaluating public transportation health benefits. Victoria Transport Policy Institute, 2016.

Loga, Wiktoria, Krzysztof Brzozowski, e Artur Ryguła. “A Method for Estimating the Occupancy Rates of Public Transport Vehicles Using Data from Weigh-In-Motion Systems.” International Conference on Transport Systems Telematics. 2016. 426-435.

London, Transport for. (2018, June 8). Bus data collection pilot. Retrieved from https://tfl.gov.uk/corporate/privacy-and-cookies/bus-data-collection-pilot

Méo, Guy Di. “Introdução ao debate sobre a metropolização. Uma chave de interpretação para compreender a organização contemporânea dos espaços geográficos.” Confins. Revue franco-brésilienne de géographie/Revista franco-brasilera de geografia (Théry, Hervé), 2008.

Mikkelsen, Lars, Radoslav Buchakchiev, Tatiana Madsen, e Hans Peter Schwefel. “Public transport occupancy estimation using WLAN probing.” Resilient Networks Design and Modeling (RNDM), 2016 8th International Workshop on. 2016. 302-308.

Ojeda, Luis Leon, Alain Y. Kibangou, e Carlos Canudas De Wit. “Adaptive Kalman filtering for multi-step ahead traffic flow prediction.” American Control Conference (ACC), 2013. 2013. 4724-4729.

Ribeiro, Renato Guimarães, “Estudo dos Deslocamentos Urbanos da Classe Média Brasileira na Região Metropolitana de Belo Horizonte”. Doctorate Thesis, COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2015.

Sindipeças. Relatório de Frota Circulante de 2018. Accessed in July 2018, de http://www.automotivebusiness.com.br/.