2016 17th IEEE International Conference on Mobile Data Management

BusesInRio: buses as mobile traffic sensors

Managing the bus GPS data in the City of Rio de Janeiro

Bruno Guberfain do Amaral, Rafael Nasser, Marco Antonio Casanova, Hélio Lopes
Departamento de Informática
Pontifícia Universidade Católica do Rio de Janeiro
Brazil
bguberfain@inf.puc-rio.br, rafael.nasser@les.inf.puc-rio.br, {casanova,lopes}@inf.puc-rio.br

***Abstract*— The main input data for a traffic management system are trajectories generated by active GPS devices installed in vehicles. In this case, such trajectories can be understood as mobile traffic sensors. A vehicle raw trajectory may be considered as a continuous data stream generated by a GPS device, installed in the vehicle, that continuously transmit its position. In some cities, such as the City of Rio de Janeiro, each bus circulating in the city has to be equipped with such device. Using this fact, the City Hall of Rio de Janeiro offers a public Web service that publishes, at about every minute, the current GPS position of all buses. However, this public data service offers only the instantaneous data, in other words, there is no historical data available. For this reason, it is necessary to develop another service that periodically queries this data and stores its entries for future processing and for useful applications, such as traffic analysis and planning. The objective of this paper is to present the *BusesInRio* tool that implements this functionality and provides access to bus trajectory data since June, 2014. The current database has approximately 2 billion GPS samples.**

Keywords—Data Stream Processing, Trajectory data management, Traffic management system.

I. INTRODUCTION

Traffic congestion in urban areas is, for sure, a serious problem, mainly because it causes an enormous economic damage. Considerable effort has been devoted to better control the traffic and to plan the mobility in a city [1]. A possible solution to diminish traffic congestion is to provide an efficient public transportation system, since it could play an important role in large cities. In [2], the authors mentioned that the main reasons for this role are because it reduces the city transportation cost, offers a service for all citizens and saves energy.

The public transportation system of the City of Rio de Janeiro is historically deficient, mainly because it is based on an old bus system. To change this state of affairs, the City Hall took some actions, such as the development of an open data project that exposes on the Web, at about every minute, the GPS instant position of all buses in the city. Although it is not a new technology, it is the first initiative to be developed in Rio. This paper describes a tool that provides access to a large dataset containing the GPS positions (more than 2 billion samples by January, 2016) of all buses that operate in the City of Rio de Janeiro since June, 2014.

We could say that the continuously transmitted bus GPS positions may play the role of mobile traffic sensors. Moreover, the raw bus trajectory could be understood as a continuous data stream acquired from the bus GPS device.

Since the City of Rio de Janeiro has a dense bus network, the acquired bus trajectories provide a useful data source for traffic analysis. In fact, the bus trajectories form a stable data source, since they traverse a fixed set of streets, at predictable regular intervals, when the traffic conditions permit. Thus, such data can be used to detect traffic abnormalities.

Contributions. This paper aims at presenting a tool, called *BusesInRio*, that collects and manages bus GPS data in the City of Rio de Janeiro. The tool also includes a graphical interface to query this huge data set. With this data, we could build traffic maps from the historic bus GPS data and investigate the traffic changes in the city over time. Although the techniques used for these purposes have been intensively investigated in the literature, this paper provides a solution for the specific case for the of the City of Rio de Janeiro, which takes advantage of several characteristics of the data.

Paper outline. Section II describes the *BusesInRio* tool. Section III presents the data acquisition and data processing algorithms. Section IV shows the graphical interface and some applications of the tool to traffic analysis. Finally, Section V concludes this work.

II. BUSESINRIO TOOL

The main objective of this paper is to describe the *BusesInRio* tool that acquires, enriches, stores and provides access to the bus GPS data in the City of Rio de Janeiro. The first step to built the tool was to implement the acquisition algorithm, in an automatic and stable way. Then, we elaborated some processing steps for route matching and speed estimation. We also had to enrich the data, for example, to identify which route a bus is serving, since approximately 23,34% of all bus trajectories (in Rio) do not have such information. We opted to store this enriched data in such a way as to achieve a good balance between the computational resources and the answering time for important queries. Finally, to give access and provide a visual analysis of the data, we developed a

2375-0324/16 $31.00 © 2016 IEEE
DOI 10.1109/MDM.2016.66

graphical Web service (available at http://www.busesinrio.com/now).

We implemented this tool using C# in the Windows Azure Cloud. We instantiate a working machine that is responsible for running the acquisition, processing and enrichment algorithms. The data is stored in such a way that they could be shared, i.e., in the Azure Blob Store. Figure 1 shows the acquisition and data storage scheme in the cloud that we use. When we need to recover the data stored in a given period to perform an analysis, we instantiate other machines to support the data processing.

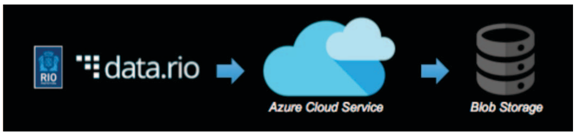

FIGURE 1. THE ACQUISITION AND DATA STORAGE SCHEME IN THE CLOUD.

In order to provide a simple description of this tool, in Section IV, we will present the algorithms for data acquisition and the processing strategies to estimate the speed of the buses. Finally, in Section V, we will describe the graphical interface of the *BusesInRio* tool.

III. DATA ACQUISITION AND PROCESSING

A. Acquisition

The data presented in this section, and used throughout this paper, has been continuously acquired from Dados Rio [3] since June, 2014. Samples from the same bus are updated at Dados Rio about every 1min30sec. Until January 2016, more than 2 billion samples were stored. Figure 2 shows the GPS sampling interval time distribution from the dataset.

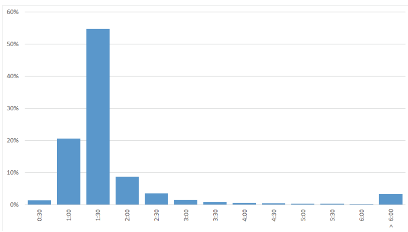

FIGURE 2. GPS SAMPLING INTERVAL TIME DISTRIBUTION OF A BUS LINE.

The public data service from Dados Rio offers only the instantaneous data, that is, no historical data is available. For this reason, we implemented a service that periodically queries this data and stores its entries for future processing. This service was implemented using the Groovy [4] language and runs 24 x 7 in a machine at the Microsoft Azure Cloud. A pseudo-code of its logic is shown in Algorithm 1. Notice that we use a delay of 40s to guarantee a complete acquisition along the day. Doing so, we could also measure the delay of the sampling time for each bus.

The log file generated by Algorithm 1 is rolled daily and contains the following information for each sample: Timestamp; Bus id (order number); Line number; Coordinates (latitude and longitude); Instant speed; Delay of timestamp, relative to Dados Rio server's date.

Algorithm 1 Pseudocode for data gathering

```
currentState ← GETREALTIMEDATA()
WAIT(40s)
loop
    newState ← GETREALTIMEDATA()
    differenceState ← newState − currentState
    if NOT-EMPTY(differenceState) then
        SAVE(differenceState)
        WAIT(40s)
        currentState ← newState
    end if
    WAIT(4s)
end loop
```

B. Route Matching

Every bus line has a set of expected routes. Matching the samples to one of these routes unveils useful information, such as the distance travelled or the time to pass through a street section (enabling the computation of the average speed per street segment). The geometry of each route, defined by a list of consecutive coordinates, is also provided by the City Hall Web service on Dados Rio. However, the data provided by the GPS service has no information about which route the bus is following at that moment. To find it, one can simply match the current bus position with the closest route, as illustrated by the thin lines of Figure 3. Unfortunately, this can lead to errors when samples are placed between two routes close to each other.

A simple solution is to find the most probable route by looking at two consecutive samples of a bus. These two consecutive positions determine a vector $\boldsymbol{V}$ (see Figure 4). These two positions are also projected on each route R_i, respecting the route direction, to create the vectors V_{Ri}. The chosen route is that that has the smallest angle with $\boldsymbol{V}$. An example of the result of this method is show on Figure 5. This problem is a simplification of what is know as *map matching*, which is the subject of specific research, such as [5]. In this paper, there was no needed to use sophisticated algorithms since the samples should be matched with a small set of possible routes.

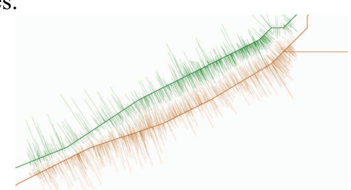

FIGURE 3. ROUTE MATCH BY CLOSEST POINT.

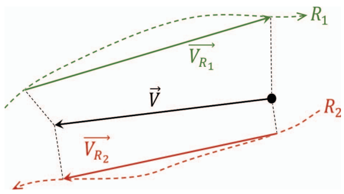

FIGURE 4. ROUTE MATCHING USING TWO SAMPLES.

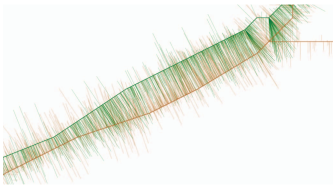

FIGURE 5. ROUTE MATCHING BY DIRECTION ANALYSIS.

Dados Rio also provides a file with the coordinates of all bus stops in the city. This information is very useful to enrich the data, since we may use it, together with the bus GPS trajectory, to identify which line and route the bus is serving.

C. *Speed estimation*

After the map-matching operation, it is possible to estimate more precisely the average speed along the bus trajectory. To do so, one can use two consecutive projected samples of a bus on a route, estimate the traversed distance and divide this estimation by the interval of time between this two samples.

IV. GRAPHICAL INTERFACE

To better understand the data and verify the acquisition process, we developed some basic visualizations using the Google Maps API [6]. We also implemented an interface that filters and facilitates accessing the data. In this section, we illustrate the use of this graphical interface.

Figure 6 shows the current position of all buses in the city at a given moment. When someone clicks on one bus icon, a pop-up appears with all details about the bus. The icons in green represent the conventional buses. The icons in blue represent express buses and the buses associated with a red icon have incomplete information, for example, their line was not informed by the Web service.

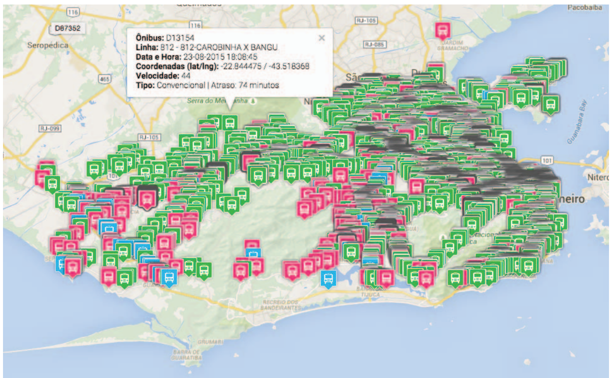

FIGURE 6. ROUTE MATCHING BY DIRECTION ANALYSIS.

Figure 7 shows the current position of all buses of Line 125, together with their routes. Figure 8 shows an equivalent image, but for the buses in Line 390. However, in this figure, there are buses that are not on their pre-defined routes, some of them are probably in their way to the garage or they have a wrong line number associated with it in the Dados Rio Web service. Figure 9 shows all bus stops associated with Line 390. Finally, Figure 10 shows the last 5 positions of a particular bus, with ID is A48081 and serving Line 415.

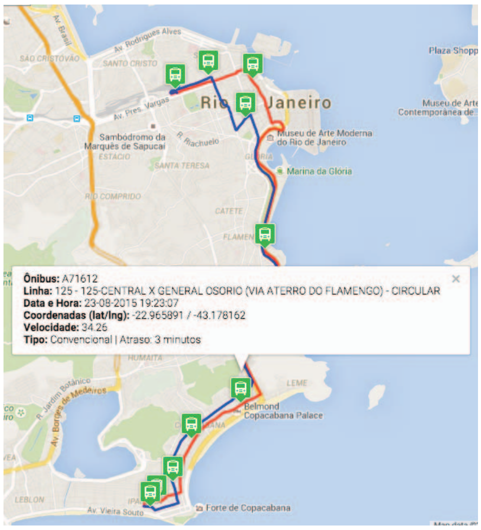

FIGURE 7. BUSES ON THE LINE 125 AND THE LINE 125'S ROUTES.

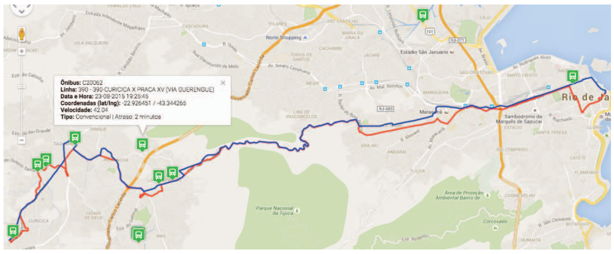

FIGURE 8. BUSES ON THE LINE 390 AND THE LINE 390'S ROUTES.

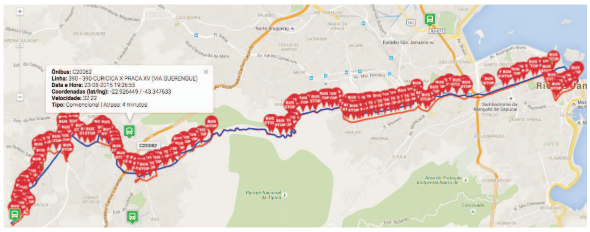

FIGURE 9. BUSES AND BUS STOPS ON THE LINE 390.

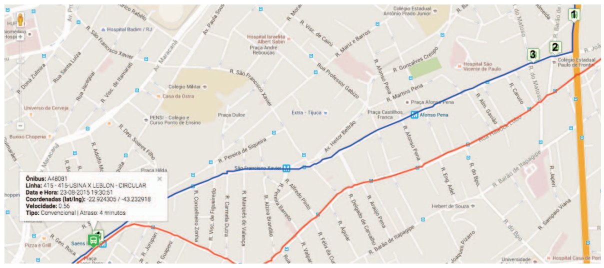

FIGURE 10. LAST 5 ACQUIRED POSITIONS OF THE BUS A48081.

The *Typical Bus Traffic* is a map that emphasizes the traffic state in a region during a given period. It is useful to spot seasonal behaviors for a particular area. In this paper, the data is computed for daily and weekly periods, each focusing on different aspects of the traffic. For example, one can use the computed daily data to observe the traffic along the hours of the day or the weekly data to spot the different behavior at a specific time on weekends and workdays. To compute the necessary information, the time dimension is discretized into time slices of fixed length, the routes are merged into shared sections and the average speed for each of these dimensions is computed. Afterwards a color scale is used to display the traffic information for a particular section at a giving time slice. The color scale implemented in this paper is based on a common scale used by popular tools, such as GoogleMaps [6] and is based on the relative speed of the route section, as shown in column Color of Table I.

TABLE I. TRAFFIC COLOR SCALE USED TO DEFINE THE TRAFFIC ON A TYPICAL BUS TRAFFIC MAP.

Label	Color	Relative speed (%)
Light		[100, 60]
Moderate		[50, 40]
Heavy		[30, 20]
Very Heavy		[10, 0]

Figure 11 illustrates an example of two maps generated for the same region, but at different time slices. Finally, Figure 12 shows the typical bus traffic map around the Botafogo neighborhood from 18:00 to 18:30 at different days of the week.

FIGURE 11. TYPICAL BUS TRAFFIC AROUND LAGOA AT 18:30 AND AT 23:30.

V. CONCLUSION

This paper summarizes some of the features of the *BusesInRio* tool, developed for the acquisition, storage and management of the bus GPS dataset made available at the Dados Rio Web site. The tool facilitates the identification of patterns, trends and anomalies in the data. Moreover, it shows how bus GPS data can be used as mobile traffic sensors.

More details about the techniques presented in this work can be found in [7].

As for future work, the authors are developing, among others, algorithms to better enrich the data and also to provide an estimation for the time of arrival of all buses in all buses stops [8].

ACKNOWLEDGMENT

The authors would like to thank CNPq, CAPES, FAPERJ and MICROSOFT AZURE FOR RESEARCH for partially supporting this work.

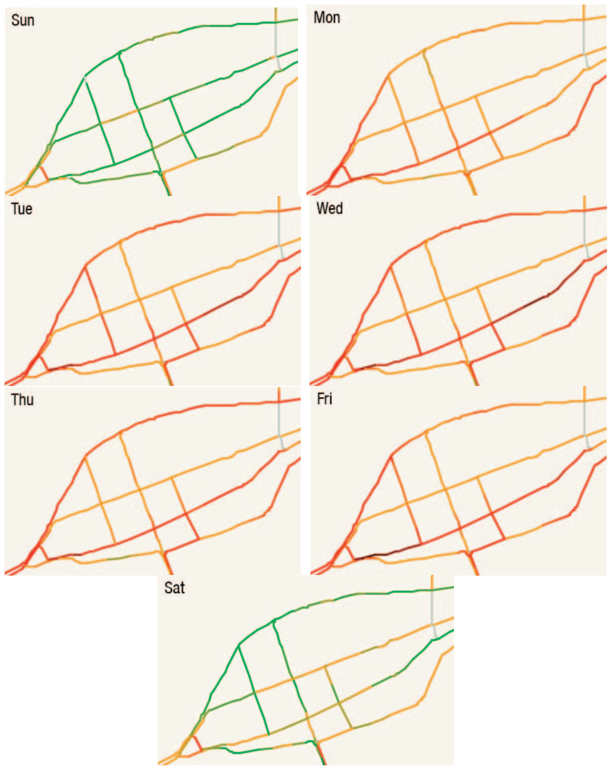

FIGURE 12. TYPICAL BUS TRAFFIC AROUND BOATAFOGO AT 18:00-18:30 IN DIFFERENT DAYS OF THE WEEK.

REFERENCES

[1] W. Shi, Q.-J. Kong, and Y. Liu, “A gps/gis integrated system for urban traffic flow analysis,” in Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on. IEEE, 2008, pp. 844–849.

[2] W. Zeng, C.-W. Fu, S. M. Arisona, A. Erath, and H. Qu, “Visualizing mobility of public transportation system,” Visualization and Computer Graphics, IEEE Transactions on, vol. 20, no. 12, pp. 1833–1842, 2014.

[3] Dados Rio, “GPS dos ônibus - Dados Rio - Prefeitura da Cidade do Rio de Janeiro,” 2014. [Online].
Available at: http://data.rio.rj.gov.br/dataset/gps-de-onibus

[4] Groovy, “Groovy language,” 2007. [Online].
Available at: http://www.groovy-lang.org/

[5] Y. Lou, C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang, “Map-matching for low-sampling-rate gps trajectories,” in Proceedings of the 17th ACM SIGSPATIAL, ser. GIS ’09. New York, NY, USA: ACM, 2009, pp. 352–361.

[6] Google Maps, “Google maps,” 2015. [Online].
Available at: https://maps.google.com/

[7] B. G. do Amaral, A visual data analysis of bus GPS data in Rio. Master Dissertation in Informatics at PUC-Rio.
(Available at http://www.inf.puc-rio.br/~lopes/bguberfain.pdf)

[8] K. L. Rodriguez, M. A. Casanova, L. A. P. P. Leme, H. Lopes, R. Nasser, B. G. do Amaral, On the Design of a Traffic Observatory Application Based on Bus Trajectories. (Accepted to ICEIS 2016)