Capacity Analysis of a City-Wide V2V Network

Daniel S. Dias and Luís Henrique M. K. Costa
GTA/COPPE, Universidade Federal do Rio de Janeiro
Email: dias@gta.ufrj.br

Marcelo Dias de Amorim
LIP6/CNRS UPMC – Sorbonne Universit
Email: marcelo.amorim@lip6.fr

***Abstract*—Passengers enjoy browsing the Internet and using social networking applications while commuting throughout the city. Nevertheless, it is not always possible to access the Internet at will. First, cellular coverage can be poor or nonexistent in some areas. Second, cellular communications may be expensive for a significant share of the population. Finding ways to offload part of the cellular traffic is desirable. In this paper, we examine the possibility of creating delay-tolerant vehicular networks in the city of Rio de Janeiro, Brazil. To this end, we rely on a large mobility dataset reporting GPS positions of 12,456 buses and 5,833 taxis during a 24-hours period. We assess the data transmission viability between vehicles and show that it is feasible to transport more than 25 TB of data in a day using only vehicles, strategically placed access points, and cheap flash drives.**

I. Introduction

Cellular networks play a central role in our daily life. People increasingly depend on their mobile data connection, either to send a text message, to read news, to see the weather forecast or to know if they will arrive late. Although cellular networks have grown dramatically, the network still does not reach the entire population [1]. The main restrictive factors are its high cost, specially limiting the amount of data that users can use, and poor coverage in determined locations.

There is a strong space-time redundancy in the content accessed in a given region, such as accessing the same news portal or watching the video that has just viralized [2]. Data sharing between users can reduce the use of cellular networks. Extrapolating, we can use data sharing through local networks to keep the population connected to the Internet during their daily commute. Thus, device-to-device communication (D2D) is a promising field for reducing the bottleneck of cellular networks and to reduce the cost of access for the users [3].

The main objective is to investigate the data transmission capacity of a D2D network composed of vehicles with IEEE 802.11 wireless interfaces acting as communicating devices. A D2D network may include vehicles, these assuming a central role, such as a proxy, storing information to distribute to mobile users. The data transportation capacity of the urban vehicles becomes relevant to the update of the information. To dimension a D2D which uses vehicles as network elements, it is necessary to evaluate the *communication capacity* of the network composed by those vehicles. The feasibility of any application running on top of this network depends on its capacity [4]. To perform this capacity analysis, we use the real GPS locations of the buses and taxis.

We use this data and statistical tools to evaluate the capacity of the network. We employ complex network concepts for the formation of a directed graph, where the vertices and edges have weights related to the vehicles, by applying clustering techniques. The analysis indicates a significant transfer capacity between areas of the city of Rio de Janeiro, reaching tens of terabytes a day at short distances (~1 km), decreasing until reaching close to a terabyte between far regions (>50 km).

978–1–5090–4671–3/16/$31.00 © 2016 IEEE

II. Buses and Taxis in Rio de Janeiro

The city of Rio de Janeiro is one of the metropolises of Brazil, having over 6 million residents. The city has more than 11,000 km of streets, where 2,576,947 vehicles (17,723 buses and 32,000 taxis) circulated in 2014.

The collected dataset is composed by real-time position data reported by 12,456 buses, updated every minute, made available by the city of Rio de Janeiro through the open data portal of the city [5]. In this work we decided to collect these data because it represents the actual shuttle of the public transport vehicles, in addition to contribute to analyze and eventually share an expressive location data collection of a large number of vehicles. The collected data is from October 7th, 2014, in a 24-hours period, and has 6,226,714 entries.

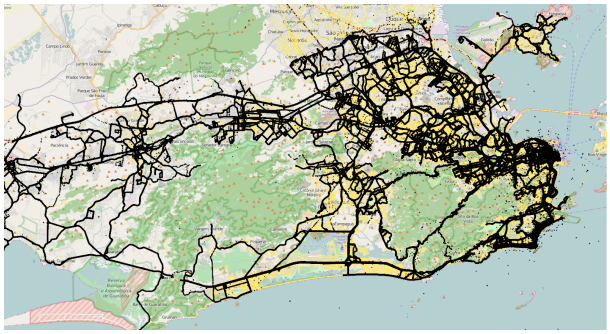


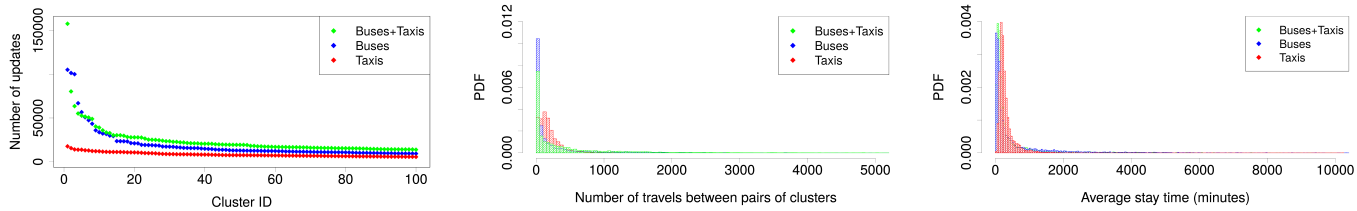
Fig. 1. Map of Rio de Janeiro with buses and taxis location data.

The taxis locations dataset was generously shared by the 99 company. Founded in 2012, it has the largest fleet of registered taxis and is present in more than 300 cities [6]. The data was collected through the 99 mobile application using the smartphones of the taxi drivers. It is composed by 3,527,300 from October 7th, 2014 and feature 24 hours of 5,833 taxi GPS positions, which represents 18% of the city's taxis.

Figure 1 shows a map of the Rio de Janeiro city. Each dot represents the raw location data collected, not considering the error of the GPS. Each black dot represents a bus location and each yellow dot represents the location of a taxi.

III. Capacity Analysis

As the computational cost to analyze all possible vehicle encounters is prohibitive, a clustering algorithm is used to



(a) Number of updates per cluster.

(b) Travels from cluster A to B.

(c) Average time a vehicle stays in a cluster.

Fig. 2. Cluster Analysis.

analyze such large dataset. The first step of our approach is to create a graph where each vertex is a region of the city. Each vertex is associated a weight according to the number of vehicles located inside the corresponding region. For the formation of regions, the city map is divided into equal-sized quadrants (squares with sides of 300 m). Then, each location record in the database is mapped to a quadrant.

A. Clustering algorithm

We assume that a vehicle is located inside a quadrant if the location obtained from GPS is within the boundaries of that quadrant. Then, the quadrants with highest concentration are classified as "clusters". Figure 2(a) shows the number of entries whose location is within a quadrant of the respective cluster, showing the number of taxis, buses, and taxis+buses.

B. Definition of a vehicle travel

A bus travel between cluster A and cluster B is defined by analyzing the timeline of a vehicle that reported its location inside cluster A and subsequently reported its location in cluster B. From the cluster A, the "source", we select the first reported time of a single vehicle. Then in cluster B ("destination") we select the first time reported later than the selected source time. This is called a travel.

Figure 2(b) displays the probability density function (PDF) of the number of travels between pairs of clusters. Most of the identified clusters have low (below 500) travels in a day. From correlating the travel and distance matrices (not shown for the sake of space), we can moreover affirm that clusters which are closer usually have more travels between them.

The stay time in a cluster is accounted from the first time a vehicle reports its location in a cluster until the last report in the same cluster. This stay time is measured only for the same travel, as measured above.

The average stay time represents the time that a vehicle stays within the operating range of a wireless access points (WAP). We assume that WAPs are installed in each cluster and have a radio range that covers the whole set of vehicles inside the cluster. To ensure a whole-cluster coverage, one of the possibilities to data transmission is to implement several access point in each cluster, enabling the buses in a cluster to communicate with at least one of the access points.

The average stay time is used to calculate the amount of data that can be sent while in range of the WAP. Figure 2(c) shows a PDF of the average time that a vehicle, traveling between a pair of clusters, remains in the range of the source cluster.

C. Data transfers

Data transportation by vehicles can be executed using different strategies. In the first type, the vehicles within the limits of the cluster are able to transfer data using the installed WAPs. To give a figure of this capacity, we assume the IEEE 802.11g standard at 54 Mbps, a conservative approach. To determine the capacity of each contact with the WAP, the total time the vehicle stays in the cluster was used. Figure 3 shows, in GB, the PDF of the total data that can be transmitted between the vehicles and 802.11g WAPs, for buses, taxis, and both combined. The tail of the taxis distribution is shorter than the others, because it is not able to send as much data. The buses+taxis distribution is lower at the 0-200 GB interval than the buses distribution, increasing the average quantity of the data transmitted in the less busy clusters.

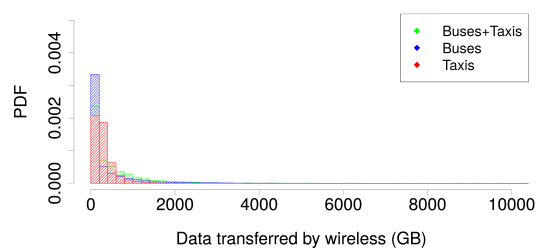


Fig. 3. Data transferred by contact between vehicles and WAPs.

The second transmission method considers the transportation of a media by the vehicles. To determine the amount of data transported, we consider the total number of vehicle travels between any two clusters. Figure 4 shows, in GB, the PDF of the data physically transported by the vehicles using physical media. This simulation considers the transfer of 4 GB of data per travel, a typical size of a flash drive or DVD. We assume that the media is embarked in the vehicle at the source cluster and disembarked at the destination. The tail of the taxis distribution is also lower than the tail of the other distributions, because of the smaller number of the vehicles, compared to the other two, and because of the lack of pattern in the movement, inducing to less travels between clusters. Taxi trajectories are less predictive than buses'. The distribution of the taxis is also relatively constant until 800 GB, decreasing its value until reach 5 TB. The taxis+buses and taxis distribution are log-normal, beginning with high values and fast decreasing until reach 20 TB for the buses and until 25 TB to buses and taxis.

Figure 5 shows the amount of data transferred by all the vehicles using wireless networks or using 4 GB media.

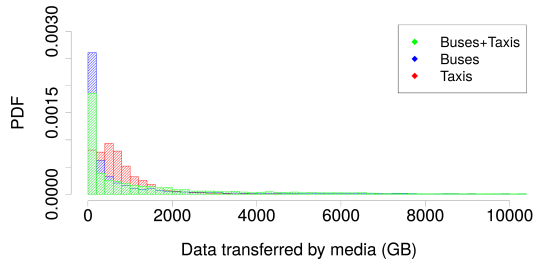


Fig. 4. Data transferred by physical media transportation.

The Y-axis represents the amount of data in TB transported between pairs of clusters while the X-axis, the distance in km. It is possible to observe a reduction in the amount of data transported as the distance increases. Although the data transmitted by wireless are predominantly higher in long range transmission (>35 km), over short ranges (<10 km) the flash drive transmission achieves higher values, reaching 25 TB. The wireless transmissions are relatively constant along the distance and most of the transmissions are below 1 TB. This almost constant behavior is explained by the average time spent at each cluster, which does not show huge variation. As those methods are independent, both can be implemented simultaneously, enabling higher data transfer capacity. For example, data transfers from cluster 3 to 20, separated by 7.2 km, can achieve 7.07 TB by wireless and 6.52 TB by media, leading to a combined 13.59 TB transfer capacity.

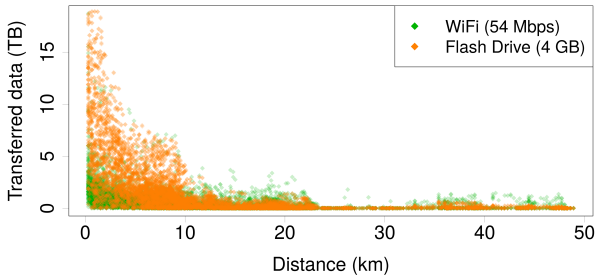


Fig. 5. Data transmission versus cluster distance.

IV. Available Vehicles Location Databases

The literature contains two types of vehicle location databases: databases generated by mathematical models and, our focus, databases collected from real-world fleet.

Location databases collected in the real world are obtained by recording vehicle positions during their motion. This type of data is many times obtained by monitoring a company fleet [7]. It is not possible yet to collect data from all the vehicles circulating in a city for several reasons, including the issue of privacy of the drivers, and because of vehicle heterogeneity: not all are equipped with a GPS and less have the equipment to record and inform their trajectory.

The real-world data available comes from bus fleets, taxi and other private company fleets. One of the first sources of data collected through the vehicle location system to be made publicly available was from the Seattle buses [8], which shows the movement of 1,200 buses, with a two-minute data collection interval. Further analysis with taxis and corporate fleets were made in San Francisco [9] and Shanghai [10]. A more detailed analysis of the mobility data available in the literature is provided by Uppoor *et al.* [7].

V. Conclusions and Future Work

The analysis of this paper aimed to verify the feasibility of physical transportation of data by urban vehicles in Rio de Janeiro. By transporting data opportunistically, it is possible to exchange significant amounts of information, in the order of tens of terabytes in a 24-hour period. In general, wireless transfers are better at higher distances while flash drive transfers are better at shorter distances. Combining them can double the data transfer. Comparing the complexity of implementation the data transport by vehicles to the complexity of creating a typical cable infrastructure or a cellular network, it is significantly easier to prepare the structure of the buses, taxis and at the location of the clusters than to install cables between the locations or install a cell tower.

As future work, new metrics can be derived from the collected data. Other concepts of complex networks should be applied to determine the formation of communities with large capacity of exchanging data [11]. Data dissemination algorithms can be applied on the traces [12], leading to a better understanding of the message exchange time and capacity.

References

[1] D. M. West, "Digital divide: Improving internet access in the developing world through affordable services and diverse content," *Brookings Institution, February*, 2015.

[2] J. Wan, D. Zhang, S. Zhao, L. T. Yang, and J. Lloret, "Context-aware vehicular cyber-physical systems with cloud support: architecture, challenges, and solutions," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 106–113, Aug 2014.

[3] D. Syrivelis, G. Iosifidis, D. Delimpasis, K. Chounos, T. Korakis, and L. Tassiulas, "Bits and coins: Supporting collaborative consumption of mobile internet," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, April 2015, pp. 2146–2154.

[4] B. Baron, M. Campista, P. Spathis, L. H. M. Costa, M. D. de Amorim, O. C. M. Duarte, G. Pujolle, and Y. Viniotis, "Virtualizing vehicular node resources: Feasibility study of virtual machine migration," *Vehicular Communications*, vol. 4, pp. 39–46, 2016.

[5] Data.rio, "Transport and mobility in the city of rio de janeiro," Website, Prefeitura do Rio de Janeiro, June 2016. [Online]. Available: data.rio/en

[6] 99Taxis, "About us," Website, 99Taxis, June 2016. [Online]. Available: http://www.99taxis.com/en/startup/

[7] S. Uppoor, O. Trullols-Cruces, and M. Fiore, "Generation and analysis of a large-scale urban vehicular mobility dataset," *IEEE Transactions on Mobile Computing*, vol. 13, no. 5, pp. 1061–1075, May 2014.

[8] J. G. Jetcheva, Y. C. Hu, S. PalChaudhuri, A. K. Saha, and D. B. Johnson, "Design and evaluation of a metropolitan area multitier wireless ad hoc network architecture," in *Mobile Computing Systems and Applications, 2003. Proceedings.*, 2003, pp. 32–43.

[9] M. A. Hoque, X. Hong, and B. Dixon, "Analysis of mobility patterns for urban taxi cabs," in *Computing, Networking and Communications (ICNC), 2012 International Conference on*, Jan 2012, pp. 756–760.

[10] M. Sede, X. Li, D. Li, M. Y. Wu, M. Li, and W. Shu, "Routing in large-scale buses ad hoc networks," in *2008 IEEE Wireless Communications and Networking Conference*, March 2008, pp. 2711–2716.

[11] D. Naboulsi and M. Fiore, "On the instantaneous topology of a large-scale urban vehicular network: the cologne case," in *ACM symposium on Mobile ad hoc networking and computing*. ACM, 2013, pp. 167–176.

[12] S. R. Cavalcanti, M. E. M. Campista, F. B. Abdesslem, L. H. M. K. Costa, M. D. de Amorim, and O. C. M. B. Duarte, "Veer: A trajectory-based peer selection algorithm for networks of vehicles," in *Med-Hoc-Net, 2010 The 9th IFIP Annual Mediterranean*, June 2010, pp. 1–8.