

Per-Vehicle Coverage in a Bus-Based General-Purpose Sensor Network

Pedro Cruz, Rodrigo Couto, Luís Henrique M. K. Costa, Anne Fladenmuller, and Marcelo Dias de Amorim

Abstract—Vehicle-based sensing is a valuable paradigm to serve third-party applications, but understanding the coverage of such networks is not trivial. For urban buses, trajectories are spatially predictable, but traffic conditions can hinder data timeliness. In this paper, we propose a metric to quantify the contribution of individual buses to the coverage of different applications. We apply this metric to bus mobility traces from Rio de Janeiro for applications of waste management, air quality, and noise monitoring, ranking buses accordingly. We show that bus contribution is strongly related to the application – 23% of buses change more than 2,000 places in different application rankings.

Index Terms—Smart cities, Internet of things, vehicular-based wireless sensor networks.

I. INTRODUCTION

Throughout their paths, buses could carry sensors and collect relevant data of a city, delivering it when reaching communication gateways [1]. The efficiency of such a system is inherently related to several parameters. For example, each bus line covers a subset of the city streets, providing partial coverage of the city. Also, buses of the same line may experience various traffic conditions, potentially causing drastic variations of their data delivery delays. Therefore, each bus is likely to have a different *contribution* in terms of city coverage, depending on the smart city application under consideration.

In the literature, several papers have investigated the coverage of bus-based mobile wireless sensor networks (MWSNs) [1]–[3]. Those studies set different coverage metrics for different classes of applications. A class of smart city applications is defined by its constraints, such as the minimum required frequency for data collection or the freshness of the collected data. These studies also show that bus-based MWSNs can disclose new applications or improve existing ones, depending on the existing bus network. These previous works identify the coverage of a bus-based network but lack an estimation of individual bus contribution in a multi-application scenario. Moreover, some works analyze the mobility of buses for smart city applications [4]–[6]. These works focus on the transport system itself. We use similar ideas to analyze the behavior of the bus network for third-party applications that can leverage bus mobility for city-wide sensing.

Quantifying individual bus contributions can help identify which buses are critical to the sensing tasks. This information allows operators to deploy a more efficient network

by, for example, dismissing some buses from sensing tasks. Therefore, it is possible to lower the global costs in terms of sensing equipment, data transmission, and data mining.

In this paper, we investigate the contribution of each bus to the coverage of a city sensed by a fleet of urban buses. We propose a metric to quantify such contribution, verifying if it is more influenced by the contributing bus or the application. We apply this metric to real GPS traces from Rio de Janeiro to rank each bus based on its importance to the applications of waste management, air quality, and noise monitoring. Our results reveal that the contribution of each bus highly depends on the application. For instance, we show that 23% of buses change more than 2,000 positions in rankings from different applications. More than this, half of the 5,856 buses have a rank that differs more than 1,000 positions for different applications. Based on these observations, one can choose buses to dismiss from sensing tasks, either to reduce costs or for maintenance reasons.

II. COVERAGE CONTRIBUTION METRIC

We denote \mathcal{B} the set of buses sensing the city and \mathcal{A} the set of applications served by the MWSN. Each bus $b_i \in \mathcal{B}$ may follow a different path or face different traffic conditions, while each application $a_j \in \mathcal{A}$ has its own requirements.

We classify smart city applications in terms of data collection requirements [7]. In this paper, we use the requirements of maximum delivery delay and minimum measurement frequency (see Section III). We define the coverage contribution of $b_i \in \mathcal{B}$ for a single application $a_j \in \mathcal{A}$ as:

$$K_{a_j}^{b_i} = \frac{C_{a_j}^{\mathcal{B}} - C_{a_j}^{\{\mathcal{B}-b_i\}}}{C_{a_j}^{\mathcal{B}}}, \quad (1)$$

where $K_{a_j}^{b_i}$ is the individual coverage contribution of bus b_i to the application a_j , $C_{a_j}^{\mathcal{B}}$ is the coverage obtained for application a_j using all the buses in \mathcal{B} (Section III shows how to obtain $C_{a_j}^{\mathcal{B}}$), and $C_{a_j}^{\{\mathcal{B}-b_i\}}$ is the coverage obtained by the MWSN when b_i is removed from the network. In a nutshell, coverage is the proportion of the city that can be sensed by the considered MWSN. Note that the contribution metric is a real number between 0 and 1.

Bus ranking. To decide which buses of the MWSN are the most or least important for a given application, we rank buses according to their contributions. The utility of a MWSN increases when it can serve several applications at the same time. In this configuration, the problem is that these applications have different requirements in terms of data

P. Cruz, Rodrigo Couto and Luís Henrique M. K. Costa are with GTA/PEE/COPPE – Universidade Federal do Rio de Janeiro e-mail: cruz,rodrigo,luish@gt.a.ufrj.br.

Anne Fladenmuller and Marcelo Dias de Amorim are with LIP6/CNRS – Sorbonne Université e-mail:anne.fladenmuller,marcelo.amorim@lip6.fr.

TABLE I
DATA REQUIREMENT AND COVERAGE FOR DIFFERENT APPLICATIONS.

Application	Requirements		Coverage (%)	Length (km)
	F_{\min} (day ⁻¹)	D_{\max} (s)		
Waste management	24	1,800	32.5	4,494
Air quality	48	300	23.8	3,292
Noise monitoring	144	300	17.4	2,416

collected. Therefore, we expect that the same bus has different contributions depending on the application. This means that some buses can be dismissed from the sensing task for some applications, but must remain in the network because they are required by some other applications. In the next section, we present the coverage model we use.

III. COVERAGE MODEL AND BUSES CONTRIBUTIONS

To compute the city coverage C thanks to the buses in \mathcal{B} , we first discretize the route map of the target city into street sections. A street section is a segment of road with a single entrance, a single exit, and a single path linking both. In a nutshell, the discretization procedure generates a graph in which vertices represent points of interest such as street intersections or steep curves. An edge exists in the graph if a road segment connects the vertices (see Section IV) [1], [3].

Let \mathcal{S} be the set of all street sections of a city and $l(s)$ be a function that returns the length of street section $s \in \mathcal{S}$. When a bus drives through a street section s , it collects data from s . Since the path of a bus is a sequence of street sections, it collects data from all those sections throughout its journey. The bus stores the data it collects until it makes a connection to a gateway. We consider that, each time a bus makes contact with a gateway, it delivers all the data collected since its last contact with a gateway. The fact that buses store and carry data until they encounter a gateway introduces some delay before that data gets delivered. We define the *delivery delay* as the time elapsed between the instant when a bus collects data and the instant when it delivers this data to a gateway. Often, data is only useful to the application if it arrives within a certain maximum delivery delay [7], which is application-specific. We denote the maximum tolerated delivery delay by D_{\max} .

It is likely that multiple buses cover the same street section. We define the *measurement frequency* of a street section $s \in \mathcal{S}$ as the number of times buses gather and deliver data within D_{\max} per unit of time. Different smart city applications might require different measurement frequencies. We denote F_{\min} as the minimum measurement frequency of an application. The values D_{\max} and F_{\min} characterize the requirements of a smart city application [7].

We say that street section s is covered if and only if, during a period of time T , buses take at least n measurements and deliver their data before D_{\max} , such that $n/T \geq F_{\min}$. We define \mathcal{S}_{a_j} as the subset of street sections that are covered by \mathcal{B} , respecting the tolerated delay and frequency requirements for the target application a_j . We can now compute the city coverage for a_j as:

TABLE II
PARAMETERS OF BUS DATASET OF RIO DE JANEIRO.

Parameter	Value
GPS coordinates before filtering	208,345,262
GPS coordinates after filtering	81,217,403
Total length of paths	38,479,109 km
Total matched sections visits	752,899,893
Total number of matched gateway contacts	38,647,268

$$C_{a_j} = \frac{\sum_{s \in \mathcal{S}_{a_j}} l(s)}{\sum_{s \in \mathcal{S}} l(s)}. \quad (2)$$

In a general-purpose MWSN, sensors may have different sensing ranges. We assume that a sensor collects data from a certain street if and only if the carrier bus passes by this street. For simplicity, we consider that streets are one-dimensional, and thus the coverage is related to street length. Of course, this coverage metric is applicable to sensors whose range is at least the street width. We further explore this metric in [8].

IV. DATA-DRIVEN ANALYSIS

We use a real-world scenario to verify the variability of buses contributions with respect to different applications. We consider three applications, namely waste management, air quality monitoring, and noise monitoring. The data requirements D_{\max} and F_{\min} of these applications are summarized in Table I, as defined by Zanella *et al.* [7]. We also extract the street sections \mathcal{S} and their lengths from a map of Rio de Janeiro provided by OpenStreetMap [9].

To obtain bus mobility traces and assess their contacts with gateways, we use the coordinates of each bus stop and a dataset containing the GPS coordinates of urban buses, both offered by the city administration of Rio de Janeiro¹. The coordinates of each bus stop are set and remain static, while the GPS coordinates of the buses are updated once a minute. We have collected data from November 1st to November 30th, 2018, obtaining GPS coordinates generated by 5,856 buses.

To acquire the bus routes in terms of street sections, we filter the GPS coordinates to remove noise and then obtain the most likely route followed by each bus using the matching function of the Open-Source Routing Machine (OSRM)². This function returns the route of each bus as a path on the graph. Every two consecutive vertices are connected by an edge; therefore, the transformation to street sections is straightforward. The main characteristics of this dataset are shown in Table II.

We need then to obtain the instants when buses collect and deliver data. Therefore, we interpolate, from the timestamps of GPS entries, the instants when buses visit each vertex in their paths. We consider that the vertices that are less than 10 m away from a bus stop are within the radio communication range of the gateway. As a result, we obtain the records of the instants when the bus has passed by the street sections, as well as the records of the moment it encountered the gateways.

¹http://dadosabertos.rio.rj.gov.br/apitransporte/apresentacao/pdf/documentacao_gps.pdf

²<http://project-osrm.org/>

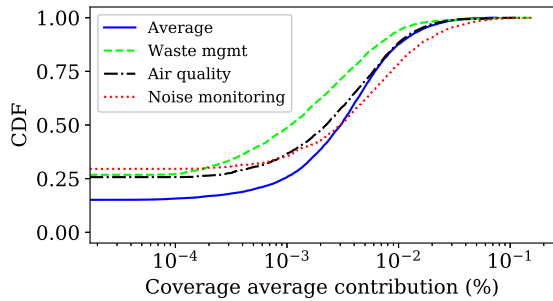


Fig. 1. CDF of the bus contributions $K_{a_j}^{b_i}$ from Rio de Janeiro.

Therefore, the delivery delay is the time elapsed between the instant a bus enters a street section (i.e., the bus visits the first edge of the street where it is about to gather data) and the instant when this bus visits a covered edge (which, we assume, provides Wi-Fi connectivity to the buses).

Meegahapola et al. and Liu et al. show that it is possible to rebuild bus routes in Singapore and London, even when GPS traces are not available [4]–[6]. This means that, with some adjustments, the methods used in this paper can be replicated to these cities. Since our coverage metric considers the delivery delay, it is expected that buses can achieve better coverage in cities with higher bus density and faster traffic.

A. Coverage contribution

Using the data obtained in the previous steps, it is possible to evaluate the coverage obtained by all the buses or by a subset of them for each considered application, as shown on Table I. More specifically, we calculate $C_{a_j}^{\{B-b_i\}}$, the coverage for an application a_j when bus b_i is not present, for every bus and every application. As a reference, the set of all buses can cover 50.0% of the city for an application with no delay or frequency restrictions. This represents a total of 6,929 km of streets. The total length of streets of Rio de Janeiro is 13,852 km. In our results, we note that there is a maximum difference of 0.04% in coverage when removing a single bus for the application of waste management, 0.02% for air quality, and 0.01% for the application of noise monitoring. In terms of the covered distance, this means a maximum of 5.8 km, 2.3 km, or 1.8 km of difference, respectively. These results set a maximum coverage loss threshold per removed bus, a useful value for scheduling maintenance, for instance.

Figure 1 illustrates the cumulative distribution of buses contributions for the applications of waste management, air quality, and noise monitoring, obtained using Equation 1 for each application. Figure 1 also illustrates the cumulative distribution of the average contribution of each bus over those three applications. The graph shows that the curve for the average contribution has much fewer buses with very low contributions than the curves for each application. It also shows that distributions depend on the applications, suggesting that applications are relevant to the contribution.

We also obtain the average speed of buses and the total coverage for each day. Considering the 30 days, the correlation between average speed and coverage is 0.834, 0.878, and

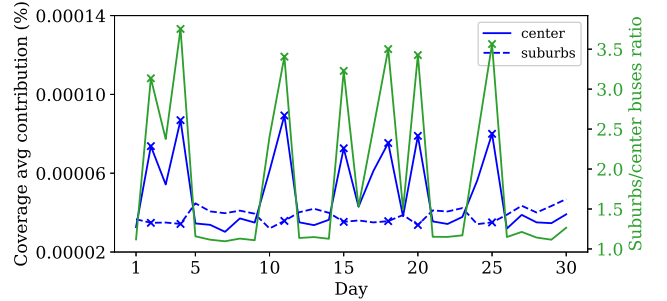


Fig. 2. Average of the daily bus contributions $K_{a_j}^{b_i}$ from Rio de Janeiro, for region and bus proportion between center and suburbs.

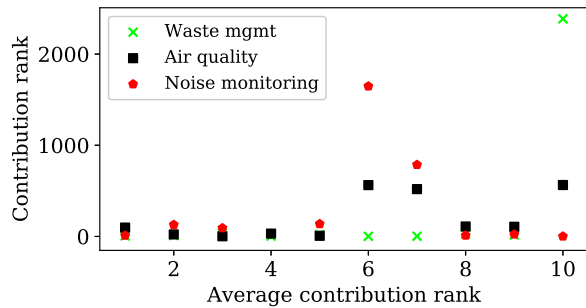
0.882 for the applications of waste management (i.e., the least restrictive one), air quality, and noise monitoring (i.e., the most restrictive one), respectively. A less restrictive application, with $F_{\min} = 1/\text{day}$ and $D_{\max} = 24 h$ would have a Pearson coefficient of 0.788. This indicates that more restrictive applications receive more influence from the traffic conditions.

City coverage is not spatially homogeneous [8]. To measure this effect to the contributions of the buses, we divide buses in two groups: center buses, which visit city center at least once, and suburb buses, which never visit the center. Figure 2 shows, for each day, the average contribution of each group, for each considered application. It also shows, in the right axis, the ratio between suburb and center buses. We indicate Sundays and holidays with an “x” mark. We note that, when the center has fewer buses, they are more important.

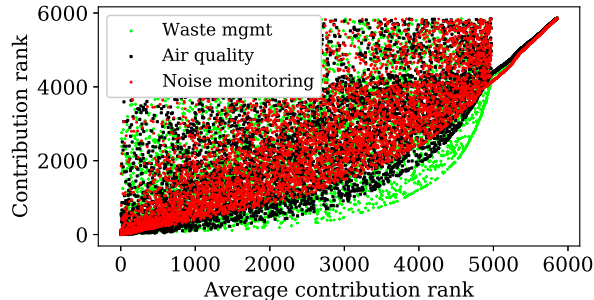
B. Contribution ranking

To quantify the impact of the influence of the application to the contribution, in the next experiment, we rank the buses for each application by ordering their coverage contributions, and also for the average of their contributions. After that, we measure the ranking difference of the same bus, for all of the considered applications, and also with respect to their average contributions.

Figure 3(a) shows, on the horizontal axis, the rank of each bus when ranked by their average contribution, for the top 10 average contributors. In the vertical axis, Figure 3(a) shows the contribution rank for the analyzed applications. Therefore, every vertical section represents the ranks of the same bus. Buses represented more to the left have a higher average contribution. We note that the most relevant buses tend to have high contributions for every application, but this tendency does not hold for all the 10 most relevant buses. The 10th bus in the average ranking is in 2, 387th, 564th, and 1st places for waste management, air quality, and noise monitoring, respectively. To study the whole dataset, Figure 3(b) extends the information in Figure 3(a) for all the buses in the dataset. It is possible to observe that, except for the buses with the lower average contribution, there is low stability of positions between rankings. This means that, in our dataset, there are not many buses that are important to every application, but there is a significant number of buses that have a low contribution to all the applications. To quantify the stability between



(a) Contribution rank as a function of average contribution rank, for different applications, for the 10 largest average contributions.



(b) Contribution rank as a function of average contribution rank, for different applications.

Fig. 3. Contribution ranking of buses from Rio de Janeiro.

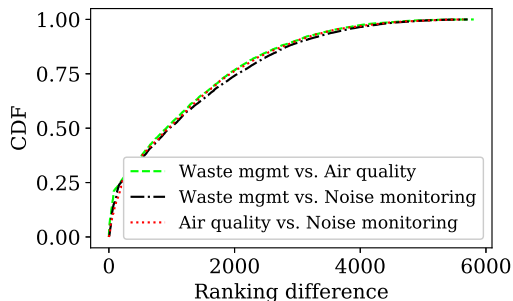


Fig. 4. Contribution ranking difference of the same bus for an application compared to its contribution rank for another application.

rankings, Table III shows the Kendall correlation coefficients of the contribution rankings of the considered applications. A Kendall coefficient of 1 means that rankings are correlated and thus independent of the application, while a coefficient of 0 means that rankings are uncorrelated. The small values of Kendall coefficients confirm that the applications have a strong influence over bus importance.

Figure 4 shows the cumulative distribution of the ranking

TABLE III
KENDALL COEFFICIENT OF BUS CONTRIBUTIONS RANKING.

Application	Kendall coefficient
Waste management \times Air quality	0.21
Waste management \times Noise monitoring	0.20
Air quality \times Noise monitoring	0.17

difference of the same bus between its position in the contribution rankings of different applications. We observe that, in every case, more than 40% of the 5,856 buses change at least 1,000 places in the ranking. This means that the proposed contribution metric is highly related to the target application. Moreover, some changes in the application requirements can also change the way buses contribute to the coverage. The proposed metric made it possible to discover and quantify the difference in each bus importance to the city coverage.

V. CONCLUSION

In this paper, we investigate the role of individual buses to the coverage of different applications in a city-wide MWSN. We propose a contribution metric to individual buses that takes into account the coverage lost for a single application when a bus is not performing sensing tasks. Using a delay-aware coverage metric, we apply the metric to GPS traces of buses in the city of Rio de Janeiro, considering the applications of waste management, air quality, and noise monitoring. We also rank the buses in terms of their city coverage contributions for each application and compare the rankings of each one. With the results, a developer can build a more efficient network, deciding which buses are crucial to each application.

As future work, we plan to use our metric to propose an optimization of the buses coverage when we have a limited number of vehicles, determining the equipment and network costs that can be saved. We also intend to improve the coverage of the network by replacing certain buses by static sensors. In both cases, the contribution of each bus can serve as the basis for heuristics for the optimizations.

ACKNOWLEDGEMENTS

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. It was also supported by CNPq, FAPERJ, and FAPESP Grant 15/24494-8.

REFERENCES

- [1] P. Cruz, R. S. Couto, L. H. M. K. Costa, A. Fladenmuller, and M. Dias de Amorim, "On the coverage of bus-based mobile sensing," *Sensors*, vol. 18, no. 6, pp. 1–12, 2018.
- [2] Y. Gao, W. Dong, K. Guo, X. Liu, Y. Chen, X. Liu, J. Bu, and C. Chen, "Mosaic: A low-cost mobile sensing system for urban air quality monitoring," in *INFOCOM*, 2016, pp. 1–9.
- [3] A. Marjovi, A. Arfire, and A. Martinoli, "High resolution air pollution maps in urban environments using mobile sensor networks," in *DCOSS*, 2015, pp. 11–20.
- [4] L. Meegahapola, T. Kandappu, K. Jayarajah, L. Akoglu, S. Xiang, and A. Misra, "Buscope: Fusing individual & aggregated mobility behavior for "live" smart city services," in *MobiSys*, 2019, pp. 41–53.
- [5] L. Meegahapola, N. Athaide, K. Jayarajah, S. Xiang, and A. Misra, "Inferring accurate bus trajectories from noisy estimated arrival time records," in *IEEE ITSC*, 2019, pp. 4517–4524.
- [6] X. Liu, Y. Zhou, and A. Rau, "Smart card data-centric replication of the multi-modal public transport system in singapore," *J. Transp. Geo.*, vol. 76, pp. 254–264, 2019.
- [7] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, 2014.
- [8] P. Cruz, R. S. Couto, L. H. M. K. Costa, A. Fladenmuller, and M. Dias de Amorim, "A delay-aware coverage metric for bus-based sensor networks," GTA,UFRJ, Tech. Rep. GTA-19-16, Jun. 2019. [Online]. Available: <https://www.gta.ufrj.br/ftp/gta/TechReports/CCC19.pdf>
- [9] OpenStreetMap contributors, "Dump retrieved from <https://download.geofabrik.de/south-america.html>," <https://www.openstreetmap.org>, 2017.