A Methodology to Assess Data Consistency in Vehicular Networks Using Participatory Sensing

Carlos Henrique de O. M. André[1], Dianne S. V. de Medeiros[2], and Miguel Elias M. Campista[1]
[1]Universidade Federal do Rio de Janeiro - PEE-COPPE/Del-Poli/GTA
[2]Universidade Federal Fluminense - PPGEET/Midiacom
{choma,miguel}@gta.ufrj.br, diannescherly@id.uff.br

***Abstract*—Today, the new paradigm of Participatory Sensing (PS) opens venue for users to contribute to a sensing system by collecting and possibly analyzing data from predetermined Regions of Interest (RoI). In this scenario, however, it is fundamental to ensure that the data collected is reliable and that anomalous data cannot influence the final result. This work proposes a methodology to identify the existence of inconsistency or unreliable data, collected by malicious or faulty sensors. The methodology is composed of three main stages. The first verifies if the sampled data is valid, while the second checks if this data is within the range of expected values. If the data is not discarded, the third step updates the system. We evaluate the proposed methodology using a dataset that records the mobility of the bus fleet of Seattle, WA – USA. We assume that users inside buses collect data regarding the vehicle speed. Results show that the methodology proposed always identify the presence of anomalous data containing maximum speed values, considering the legitimate values of average and standard deviation speeds. Even manipulating the system, the impact on the anomaly detection becomes relevant only in extreme cases.**

I. Introduction

Recently, the mobile communications industry has experienced a great technological evolution as a consequence, at least partially, of the new Internet of Things (IoT) paradigm [1]. In this context, the high availability of mobile devices is changing users' role from simple data consumers to data sources for many IoT applications. Users can sense and send data to a central infrastructure typically on the Internet. This central infrastructure is responsible for analyzing and transforming the received data into useful information. The participation of multiple distributed users as data sources gives room to another recent paradigm known as Participatory Sensing (PS) [2]. Applications relying on PS benefit from abundant data gathered from larger areas at relative low cost. An example of this approach is the operation of Waze, which offers information about traffic conditions using users' contributions as input data [3].

Several sensing systems have been proposed in the literature with different goals. For instance, Cruz et al. [4] propose a system that leverages bus mobility to enlarge the monitored area using several micro weather stations, composed by meteorological sensors embedded on the urban buses. Mohan et al. [5] propose Nericell, which uses sensors embedded on smartphones, such as the accelerometer, microphone, and GPS, along with the GSM radio signal, to detect potholes, bumps, noise and changes in the movement of the vehicle. The sensed data can be then used to determine the traffic and road conditions. Zhou et al. [6] propose a system based on collaborative effort to predict the bus position over time on a trip, using the power between the cellular towers and mobile devices. All these works can be categorized under the PS paradigm. None of them, however, concerns about the contamination of the sensed information with inconsistent data.

This work proposes a methodology to evaluate the consistency of data collected by vehicular users contributing to a PS system. The proposed methodology is composed of three main stages: sample validation, anomaly detection, and system update. The overall idea is to detect and discard possible anomalous data, guaranteeing the accuracy of the processed information. The anomalous data can be produced as a consequence of sensor failures, errors during data transfer, or they can be even injected into the system by malicious users. In any case, an important characteristic to detect inconsistent data is the variability of the values measured compared with an expected value. The larger the variation, the smaller the capacity of identifying inconsistency. Hence, this work quantifies the increase on the volume of gathered data in a Region of Interest (RoI) and evaluates its influence on the validity and accuracy of the consolidated information. To this end, we use an entire day of the bus mobility dataset of the city of Seattle [7], assuming that the sensed data is the bus speed. Our proposed methodology is evaluated considering the existence of three RoIs, divided according to the average speed of the buses. We verify the impact of inconsistent samples on each scenario and we consider the existence of users whose sensors contribute with permanent or intermittent inconsistent data. Results show that it is possible to always identify the presence of anomalous data containing maximum speed values in all regions considering the legitimate values of average and standard deviation speeds, even using a very simple detection method. Even manipulating the average and the standard deviation values, the impact on the anomaly detection is not important even adding 100% of false samples.

This paper is organized as follows. Section II overviews participatory sensing. Section III introduces the proposed methodology to assess data consistency. The dataset and the main route used in our analysis are characterized in Section IV and V, respectively. Section VI presents our results. Lastly, Section VII concludes this work and presents next directions.

II. Participatory Sensing

The evolution of the mobile technology and the miniaturization of components give room to an increasing number of built-in sensors on vehicles and personal devices such as smartphones. Nowadays, sensors such as accelerometer, GPS, gyroscope, compass, microphone and camera are widespread and easily found in several devices. Hence, the sensing capacity of each device increases, allowing local knowledge acquisition and context comprehension. When these devices are collaboratively used, together extracting information from their surroundings, it is possible to build a scalable and low-cost sensing system. After suitable analysis, the data gathered by the devices can collectivity offer opportunities to develop new applications [8]. Currently, sensor networks based on the participation of users and their own devices are the focus of the Participatory Sensing (PS) paradigm [9], [10].

The participatory sensing enables to enrich the knowledge from an RoI at a low cost using data from users. Such participation, however, brings challenges due to the data heterogeneity and volume. Users must process and interpret different types of data collected along an entire trip and transfer this data to other nodes, e.g., to a gateway, during possibly short contact times. In vehicular networking, the sensed data can be transferred from On Board Units (OBUs) to RoadSide Units (RSUs) via access networks using IEEE 802.11, e.g., IEEE 802.11p, or another mobile technology, e.g., 4G. The last one, however, can add costs to users.

III. Data Consistency in Vehicular Networks

This section presents the methodology proposed to detect anomalous data in vehicular networks using participatory sensing. We also discuss user profiles that can possibly inject anomalous data into the system. The terms inconsistent and anomalous data are used interchangeably. We stress the broader sense of data inconsistency, which can be a consequence of faulty sensors and transmissions issues, as well as a consequence of malicious action.

A. Proposed methodology

The proposed methodology is composed of three main stages: sample validation, anomaly detection, and system update. We highlight that our methodology is general enough to yield other sensor types. In addition, it can be executed on an Internet server, or locally, on a sensing device. Hybrid configurations are possible, meaning that the methodology is partially executed on devices and partially on a cloud server, for instance. Figure 1 depicts the methodology operation.

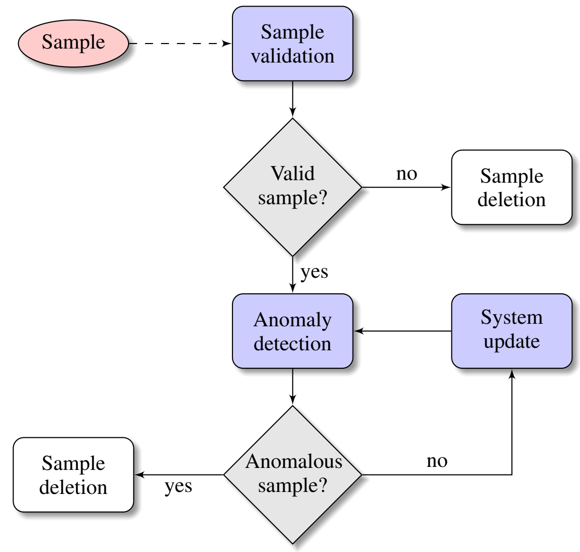


Fig. 1. Proposed data consistency methodology for PS in vehicular networks.

Sample validation**:** samples containing invalid values for the sensed data are discarded. Hence, we avoid the waste of storage and processing resources with samples that do not contribute to the participatory system. Considering a sample x_i composed by a vector of properties $\langle p_{i,1} \ldots, p_{i,n} \rangle$, this sample is considered valid only if each one of its properties falls within a predefined known range, i.e., $p_{i,j} \in \mathcal{D}_j$. In this work, the GPS location and the value of the bus speed are used as sample properties. We assume that the location must be within the territory of the city where the data was collected. The absolute speed is always positive and must be under a threshold, which can vary from one scenario to another. This threshold can take into account several factors, such as the highest speed achievable by a modern vehicle, the mobility in urban or rural areas, the presence of highways, etc.

Anomaly detection**:** anomalous data are discarded. To this end, we assume the existence of a system able to judge whether a valid sample is within an expected range of values. The notion of expected value can be built based on the history of received values. The methodology is generic enough to use different detection systems. We evaluate only the property related to the value of the sensed data. In this work, the proposed methodology uses a simple detection mechanism. Consider x_i the sample analyzed, a sample is considered as anomalous if it falls outside the range $[\mu - 3\sigma, \mu + 3\sigma]$, where μ is the average of the last received values and σ is the corresponding standard deviation. As such, if $x_i \notin [\mu - 3\sigma, \mu + 3\sigma]$, x_i is discarded.

System update**:** if sample x_i is considered as valid and consistent, it is used to update the anomaly detection system. Note that the update process depends on the detection method used. In this work, a new valid and consistent sample is used to recalculate the average, μ, and the standard deviation, σ, used in the anomaly detection stage.

B. User profiles

We assume that users can send invalid data; valid and non-anomalous data, indeed contributing to the system; and valid, but anomalous data. The anomalies can result from transient or permanent failures of the sensors embedded into users' devices. The data can be compromised during the transfer to the data concentrator even if the user is not malicious, as a consequence of problems on data formatting, hardware malfunctioning, or even data processing. Malicious users, in turn, can produce inconsistent data, injecting them into the sensing system with the goal of manipulating the final result. This manipulation can bring advantages from the application point of view or can be used to simply harm the system operation.

In this work, we aim at analyzing the performance of our proposed methodology to assess data consistency in vehicular networks, assuming that the collected data is vehicular speeds. We consider as inconsistent, samples injected with speeds higher than the expected value for a given Region of Interest (RoI). Even though inconsistent data can also be injected by failures, as mentioned earlier, malicious users could be interested on attacking an application for vehicular traffic monitoring, such as Waze. If a malicious user could increase the expected speed of a street, then he would be able to locally interfere on the flow of vehicles, provoking for instance traffic jams. In our scenario, the average speed is not high and, therefore, low speeds that could be also considered anomalies were not found in our analysis (Section VI). Note, however, that the proposed methodology is agnostic to this characteristic. Yet, the impact of an inconsistent data is proportional to the difference between the measured and the expected values. Hence, if the value of the anomalous data falls outside the range of expected values, it will be discarded anyway.

We assume the existence of two types of users that can introduce inconsistent data. The first one introduces samples with *permanent inconsistencies*. This means that all samples injected present high speed values. Such users are called *users with permanent inconsistencies*. The second type of users introduces samples that interchanges high and average speed values, i.e., *intermittent inconsistencies*. These users are called *users with intermittent inconsistencies*. Both types can inject a different number of data samples into the system during a trip. In our analysis, however, we do not combine both types of users, i.e., either we have users with permanent or intermittent inconsistencies. The value of the anomalous speed, $s_{anomalous}$, introduced by each contribution, can be within the range $0 \leq s_{anomalous} \leq s_{max}$, where s_{max} is the maximum allowed speed in the city. The number of injected samples varies gradually between 0 and 100% of the maximum number of existing samples in the system during the analyzed one-hour interval. The results obtained for both types of user profiles are presented in Section VI.

IV. Datasets

The performance of the proposed methodology is evaluated using the dataset of the bus fleet of the city of Seattle, WA – USA (Ad Hoc City dataset [7]). This dataset contains the real position of all city buses formatted as 6-tuple entries $\langle d, t, id_{bus}, id_{route}, x, y\rangle$. Each tuple describes the Cartesian coordinates x, y (in feet) of a bus id_{bus} in route id_{route} on date d and at time t. The dataset was collected between $31/10/2001$ and $02/12/2001$, totaling 125 MB.

We analyze a typical Wednesday, $31/10/2001$, which includes information about the mobility of 236 bus routes. The bus speed is not included in the original dataset and thus we compute and attach it to each entry. To accomplish that, we consider the bus position and the time interval between two consecutive entries of the same bus. Thus, taking into account two consecutive entries of bus id_{bus} at instants t_1 and t_2 on the same day d, respectively, $\langle d, t_1, id_{bus}, id_{route}, x_1, y_1\rangle$ and $\langle d, t_2, id_{bus}, id_{route}, x_2, y_2\rangle$, the bus speed is computed as $\frac{\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}}{t_2-t_1}$. This speed is attached to the second entry, the one of instant t_2, and the same procedure is repeated considering instants t_2 and t_3 and so on until the last sample for the bus id_{bus}. The dataset is totally enriched after computing the speeds of all buses. We assume that all entries in the dataset represent users' contributions to the participatory sensing system.

The maximum speed of buses in the metropolitan area of Seattle is approximately equal to 56 km/h [11]. After computing the speed of buses, we found some unreal values that are discarded at the first stage of the proposed methodology, i.e., at the *sample validation* stage. To this end, we execute two evaluations. Firstly, we verify if the speed is unreal considering the bus speed s_i of a given entry. If s_i in km/h is not within the range $[0, s_{max}]$, we do not consider the corresponding entry. In this work, $s_{max} = 110$ km/h, which is the maximum allowed speed in the city of Seattle [11]. Secondly, we verify the GPS location of each 6-tuple entry and we delete all the entries falling out of the city of Seattle.

In the studied dataset, there exist $376{,}491$ entries, with an average of $1{,}595$ entries per route. We chose to analyze the route $id_{route} = 007$ (Route 007), characterized in the next section, because this route is the one with the highest number of entries. The sample validation stage of Route 007 discarded 55 entries, which can be neglected as they represent only 0.35% of the total number of entries of this route.

V. Route 007 Characterization

In this section, we detail the characteristics of the chosen route to facilitate the analysis of the proposed methodology. Route 007 crosses the city, starting at *Prentice Street/Rainier Beach*, located at the south of Seattle, and finishing at *Downtown Seattle*, at the north. The route length is approximately 15.45 km and the average duration of a complete trip on weekdays takes 44 minutes [12]. On the analyzed day, there are 51 buses with an average of 305 samples per bus.

A. Average speed of buses

Considering the entire Route 007, the average speed during the whole day is 11.90 km/h. We detail this result by dividing the route into small areas of size 50×50 m^2. For each small area, we compute the average speed of the buses during the same day. Figure 2 shows the average speed obtained for each area. In this figure, the redder the area, the higher the average speed. In opposition, the bluer the area, the lower the average speed. We observe that the average speed at some points of Route 007 during the day can reach up to 60 km/h. This means that there is not significant traffic jam in these points, as the value coincides with the highest allowed speed for buses in the city of Seattle [11]. We highlight in Figure 2 the division of

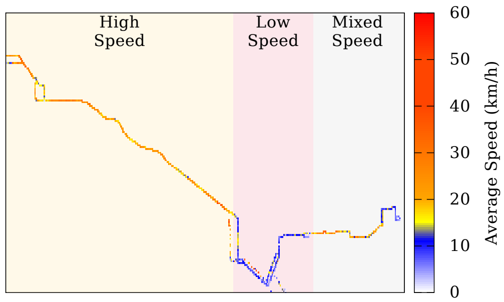


Fig. 2. Route 007 colored with the average speed of buses. We divide the route on three different regions, i.e., high, low and mixed, according to the average speed of the buses.

TABLE I
CHARACTERISTICS OF THE REGIONS DEFINED BASED ON THE AVERAGE BUS SPEED.

Name of the region	Total # of samples	Total # of buses	Average speed (km/h)
High Speed	6,809	50	17.17
Low Speed	7,172	50	7.18
Mixed Speed	1,565	32	10.63

the route in three regions based on the average speed of buses: *high speed*, *low speed*, and *mixed speed*. Table I presents the nomenclature adopted for each region and the corresponding total number of samples, total number of buses and their average speed in each region. The division in regions allows us to separate the route onto segments with similar characteristics.

B. Total number of buses and samples

The number of buses contributing during the day varies between 3 and 31 buses, depending on the analyzed hour. Nevertheless, the number of samples obtained at each hour is not proportional to the number or buses participating on the sensing task. For instance, the highest number of buses available to collect data during an one-hour interval is 31 buses, which are able to collect 998 samples. The highest number of samples, however, is 1,240, which are obtained with 19 buses only. The lowest number of samples is 265, obtained with 3 buses. The average number of buses per hour during the day is 19 and the average number of samples per hour is 777. The average number of samples per bus per hour, in turn, is equal to 59. It is worth mentioning that there is a tradeoff between volume and data representativeness. On the one hand, the amount of data cannot be very large, so that the volume transfered does not exceed the capacity of the network or the data concentrator. On the other hand, the volume cannot be very small to the point it loses representativeness.

VI. ANALYSIS OF THE PROPOSED METHODOLOGY

In this section, we aim at verifying the viability of anomalous data detection. To accomplish that, we use the two user profiles defined in Section III and we simulate scenarios with anomalous users injecting different number of samples into the system. Note that, although the analysis concerns speed samples, our methodology is generic enough to be reused in other scenarios, considering other types of sensors, e.g., pollution, luminosity, and pressure.

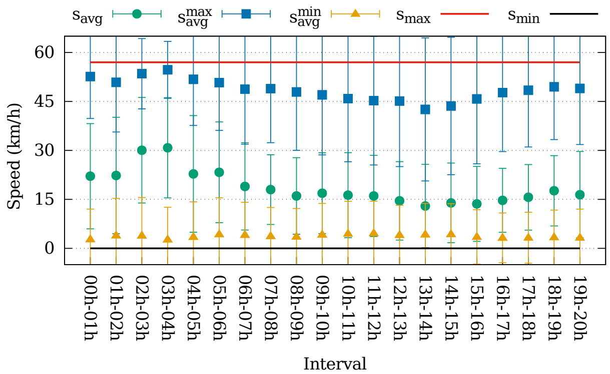
A. Impact of the anomalous samples on average speeds

In this section, we investigate the number of anomalous samples needed to change the final result. We assume that each anomalous user can inject samples to modify the average speed of the region, increasing or decreasing the real average value. On the one hand, to reduce the average speed of the region, the user must contribute with a speed value lower than the average, and the lowest possible one is 0 km/h. On the other hand, to increase the average speed of the region, the user must introduce samples with higher speed values, and the highest possible one is 57 km/h in the evaluated scenario. Both attempts to change the average speed can cause problems to the system if they are not detected. This corroborates the need for the *anomaly detection* stage of the proposed methodology.

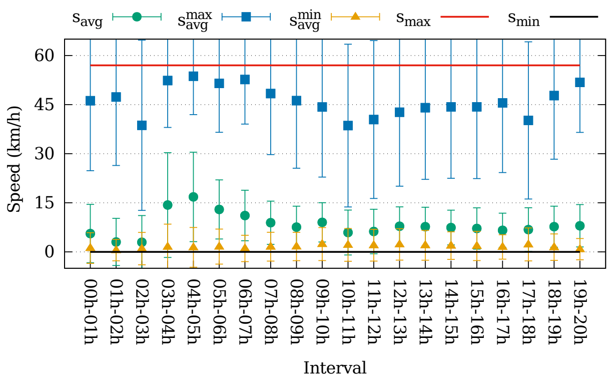
Figure 3 shows the average speed of buses for each period of 1 hour. The points $\mathrm{s_{avg}}$ show the results when the system is free from anomalous data. The points $\mathrm{s_{avg}^{max}}$, in turn, represent the average speed considering anomalous samples containing maximum speeds. Finally, the points $\mathrm{s_{avg}^{min}}$ represent the average speed considering anomalous samples containing minimum speeds. These last two results were obtained using 777 additional anomalous samples, as this is the average number of samples per hour (Section V-B). The vertical bars represent the average value $\pm 3\sigma$.

We observe that the anomalous samples carrying minimum speeds ($\mathrm{s_{min}}$), if used, would be able to decrease the average speed of each region. Considering that these anomalous samples are seldom detectable using the simple approach of $\mathrm{s_{avg}} - 3\sigma$, the impact would not create a new problem. We would only have more false positives and negatives. This happens because the average speed of each region is not high, so that it would be necessary the existence of anomalous samples with negative speeds, which are not valid, to sufficiently decrease the average speed of each region. Alternatively, it would be necessary the utilization of more sophisticated methods to detect as anomaly the downward shifting of the average speed.

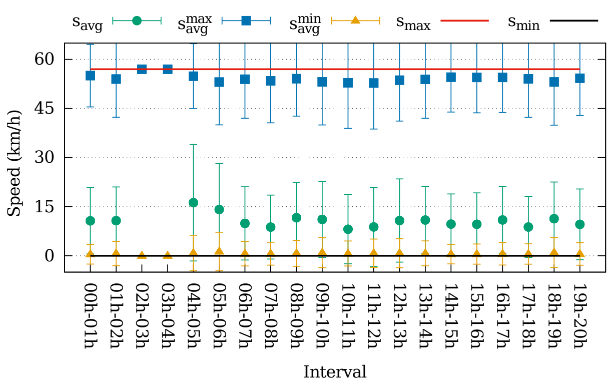
Considering now the samples containing the maximum speed ($\mathrm{s_{max}}$), all of them are already detectable in all regions using the method $\mathrm{s_{avg}} + 3\sigma$. In opposition to the minimum speed samples, the maximum speed samples are substantially higher than the average speed. If, however, these samples are used to compute the average speed ($\mathrm{s_{avg}^{max}}$), the final result would completely change. All anomalous samples would fall within the interval of expectations and almost all the legitimate samples would be considered anomalous. Of course, the impact depends on the scenario and on the number of anomalous samples, but the observed trend would still be true. The remaining of this paper focuses on the detection of anomalous maximum speed samples, using the previously



(a) High Speed region.



(b) Low Speed region.



(c) Mixed Speed region.

Fig. 3. Impact of the anomalous samples on the sensing system on each region considering the injection of 777 anomalous samples per hour.

defined user profiles. Note that in the interval 2-4h the Mixed Speed region does not have samples (Figure 3(c)).

B. Impact of the anomalous samples on the detection method

We conduct 21 simulation runs and we assume that the system is initially free from anomalous data. At each run, we add to the legitimate samples of each hour, 5% of anomalous samples to recompute the average and the standard deviation. These 5% are proportional to the highest original number of samples per hour in each region. At the last simulation run, we have 100% of additional anomalous samples. The samples introduced by users with permanent anomaly have the highest possible value of speed allowed for buses in the city of Seattle. Users with intermittent anomaly, in turn, contribute with samples interchanging between the average speed of the region and the maximum speed allowed. The number of anomalous samples, i.e., the highest number of anomalous samples per hour, and the speed values are summarized in Table II. The average speed can change depending on the region and on the current one-hour interval.

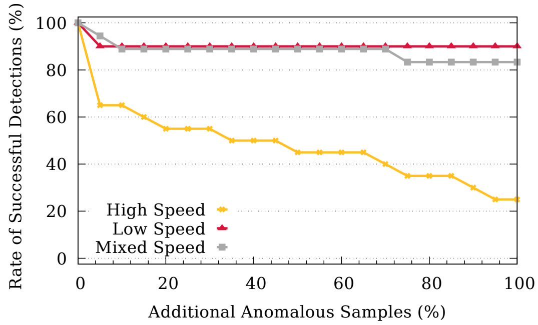
TABLE II
CHARACTERISTICS OF ANOMALOUS DATA INTRODUCED IN THE ANALYSIS OF THE PROPOSED METHODOLOGY.

		Value of the anomalous speed (km/h)	
Name of the region	**# of anomalous samples**	**Permanent anomaly**	**Intermittent anomaly**
High Speed	580	57	Average and 57
Low Speed	660	57	Average and 57
Mixed Speed	120	57	Average and 57

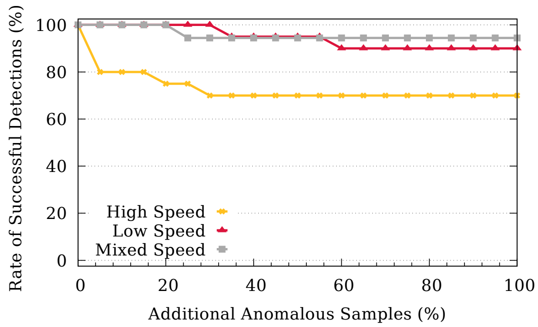
Figure 4(a) shows the rate of successful detections as a function of the number of anomalous samples injected into the system. The Y-axis represents the fraction of hours of the analyzed one-day period at which the anomalies are successfully detected. Each curve in this figure represents one of the regions previously defined. Note that detecting an anomalous maximum speed sample depends exclusively on the average and on the standard deviation computed in the current hour. If $s_{max} > s_{avg}^{max} + 3\sigma$, we guarantee that the anomalous sample is detected.

We observe that the rate of successful detections of the high speed region drops down faster than the others. This happens because the average speed is closer to the anomalous maximum speed. Hence, with few anomalous samples, we have already a reduction on anomalous detections. In all regions, the system is able to detect all anomalous samples using only the legitimate samples to compute the average and the standard deviation values, i.e., when $x = 0\%$. Nevertheless, even though the low speed region presents the best result, as expected, it quickly drops down the rate of successful detections to 90%, when the average and standard deviation values start to become manipulated using the additional 5% of anomalous samples. This reduction at the beginning is a consequence of the high average and standard deviation values legitimately found during the interval between 3-5h. The mixed region exhibits an intermediate behavior considering the previous two regions. A remarkable observation is that the rate of successful detection never reaches 0%, even when the average and the standard deviation values are highly manipulated, i.e., when $x = 100\%$.

Figure 4(b) shows a similar result, but for intermittent anomalies. As expected, in general, a higher number of anomalous samples is needed to reduce the rate of successful detections. This is because half anomalous samples has the legitimate average speed value. A remarkable difference, however, between Figures 4(a) and 4(b) is that even in the presence of 100% additional anomalous samples, the rate of



(a) Users with permanent inconsistencies.



(b) Users with intermittent inconsistencies.

Fig. 4. Influence of the additional anomalous samples, used to recompute the average speed at each time, on the rate of successful detections in each region using the simple detection method employed in the proposed methodology.

successful detections never reaches values lower than 70%.

C. Impact of anomalous samples on the network load

This analysis considers that the data is transmitted using the IEEE 802.11p standard, which is used in vehicular networks to provide communications between OBUs, and between OBUs and RSUs. The idea is to evaluate the impact of the additional samples on the network throughput. The RSUs can be connected to a central controller to have a complete view of the network. The physical layer of the IEEE 802.11p operates at a maximum transmission rate of 27 Mb/s, and minimum of 3 Mb/s, using an operating frequency of 5.890 GHz and 10 MHz bandwidth [13].

We consider that each sample has 480 Bytes [7] and the average number of samples injected per hour is 777, totaling an additional load of 372.96 kBytes. The additional anomalous samples may introduce network issues as the total load would reach 745.92 kBytes for 100% extra samples. In one hour, the aggregated network throughput would need to be 12.78 Mb/s, which may not be possible depending on the network configuration.

VII. Conclusions

This work proposed a methodology to assess data consistency in vehicular networks using a participatory sensing system. An anomaly can be a simple inconsistency or untrusted data, which can lead to false conclusions. We presented two user profiles that can produce and inject anomalous data into the system, the users with permanent and intermittent inconsistencies; and different regions separated according to the average speed of vehicles. Our analyses considering each type of user showed that permanent inconsistencies are easier and quicker to detect. The intermittent one, however, needs more samples to be detected because it interleaves real and unreal measurements. Results show that it is possible to always identify the presence of anomalous data containing maximum speed values in all regions considering the legitimate values of average and standard deviation speeds, even using a simple detection method. Even manipulating the average and the standard deviation values, the impact on the anomaly detection is not important even adding 100% of false samples. The additional volume of samples, depending on the network configuration, may result on a high network load.

As future work, we plan to evaluate our proposed methodology using all routes in the city of Seattle, other user profiles, and new scenarios. We would like also to improve our detection system using more sophisticated methods.

Acknowledgments

The authors thank CAPES, CNPq, and FAPERJ.

References

[1] I. I. Reports, "ITU Strategy and Policy Unit (SPU)," International Telecommunication Union (ITU), Tech. Rep., 2005.
[2] Y. Xiao, P. Simoens, P. Pillai, K. Ha, and M. Satyanarayanan, "Lowering the barriers to large-scale mobile crowdsensing," in *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*. ACM, 2013, p. 9.
[3] V. Ribeiro Neto, D. S. V. Medeiros, and M. E. M. Campista, "Analysis of mobile user behavior in vehicular social networks," in *International Conference Network of the Future (NOF 2016)*, 2016, pp. 1–5.
[4] P. Cruz, R. S. Couto, and L. H. M. Costa, "An algorithm for sink positioning in bus-assisted smart city sensing," *Future Generation Computer Systems*, 2017, (In press, Accepted Manuscript).
[5] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: rich monitoring of road and traffic conditions using mobile smartphones," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 2008, pp. 323–336.
[6] P. Zhou, Y. Zheng, and M. Li, "How long to wait?: predicting bus arrival time with mobile phone based participatory sensing," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACM, 2012, pp. 379–392.
[7] J. G. Jetcheva, Y.-C. Hu, S. PalChaudhuri, A. K. Saha, and D. B. Johnson, "CRAWDAD dataset rice/ad_hoc_city (v. 2003-09-11)," http://crawdad.org/rice/ad_hoc_city/20030911/bus_mobility, Sep. 2003, traceset: bus_mobility.
[8] P. C. F. Melo *et al.*, "Csvm: Uma plataforma para crowdsensing móvel dirigida por modelos em tempo de execução," 2014.
[9] W. Z. Khan, Y. Xiang, M. Y. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 402–427, First 2013.
[10] J. A. Burke, D. Estrin, M. Hansen, A. Parker, N. Ramanathan, S. Reddy, and M. B. Srivastava, "Participatory sensing," *Center for Embedded Network Sensing*, 2006.
[11] "Seattle department of transportation," http://www.seattle.gov/transportation/sdotfaqs.htm, 2017, acessado: 20/07/2017.
[12] K. C. Metro, "Route 7 - king country," http://kingcounty.gov/depts/transportation/metro/schedules-maps/007.aspx#weekday, 2017, acessado: 13/06/2017.
[13] D. Jiang and L. Delgrossi, "Ieee 802.11 p: Towards an international standard for wireless access in vehicular environments," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*. IEEE, 2008, pp. 2036–2040.