CISTER

Research Centre in
Real-Time & Embedded
Computing Systems

Conference Paper

Experimental Evaluation of Urban Points-of-Interest as Predictors of I2V 802.11 Data Transfers

Pedro Miguel Santos*

Luís M. Sousa

Ana Aguiar

*CISTER Research Centre

CISTER-TR-190901

2019/10/14

Experimental Evaluation of Urban Points-of-Interest as Predictors of I2V 802.11 Data Transfers

Pedro Miguel Santos*, Luís M. Sousa, Ana Aguiar

*CISTER Research Centre

Polytechnic Institute of Porto (ISEP P.Porto)

Rua Dr. António Bernardino de Almeida, 431

4200-072 Porto

Portugal

Tel.: +351.22.8340509, Fax: +351.22.8321159

E-mail: pss@isep.ipp.pt

https://www.cister-labs.pt

Abstract

Smart Cities will leverage the Internet-of-Things(IoT) paradigm to enable cyber-physical loops over urbanprocesses. Vehicular backhauls contribute to IoT platforms byallowing sensor/actuator nodes near roads to explore opportunisticconnections to passing vehicles when other communicationbackhauls are unavailable. A placement process of nodes thatincludes vehicular networks as a connectivity backhaul requiresestimates of infrastructure-to-vehicle (I2V) wireless service atpotential deployment sites. However, carrying out I2V measurementcampaigns at all potential locations can be very expensive;so, predictive models are necessary. To this end, qualitativecharacteristics of a potential site, such as infrastructural pointsof-interest (POI) relating to traffic (i.e., traffic lights, crosswalks)and fleet activities (i.e., bus stops, garbage bins) can inform aboutthe vehicles 19 mobility patterns and quality of the I2V service. Inthis paper, we show the contribution of POI (and site-specificinformation) to I2V transfers, leveraging a real-world dataset ofgeo-referenced I2V WiFi link measurements in urban settings.We present the distributions of throughput with respect todistance per POI class and site, and apply exponential regressionto obtain practical throughput/distance models. We then usethese models to compare I2V transfer estimation methodologieswith different levels of POI-specific data and data resolution. Weobserve that I2V transfer estimate accuracy can improve froman average over-estimation of 18.3% with respect to measuredvalues, if site or POI-specific information metrics are not used,to 9.3% in case such information is used.

Experimental Evaluation of Urban Points-of-Interest as Predictors of I2V 802.11 Data Transfers

Pedro M. Santos*†, Luís M. Sousa* and Ana Aguiar*‡
*Faculdade de Engenharia, Universidade do Porto, †CISTER, Instituto Politécnico do Porto, ‡Instituto de Telecomunicações
{pmsantos, lm.sousa, anaa}@fe.up.pt

***Abstract*—Smart Cities will leverage the Internet-of-Things (IoT) paradigm to enable cyber-physical loops over urban processes. Vehicular backhauls contribute to IoT platforms by allowing sensor/actuator nodes near roads to explore opportunistic connections to passing vehicles when other communication backhauls are unavailable. A placement process of nodes that includes vehicular networks as a connectivity backhaul requires estimates of infrastructure-to-vehicle (I2V) wireless service at potential deployment sites. However, carrying out I2V measurement campaigns at all potential locations can be very expensive; so, predictive models are necessary. To this end, qualitative characteristics of a potential site, such as infrastructural points-of-interest (POI) relating to traffic (i.e., traffic lights, crosswalks) and fleet activities (i.e., bus stops, garbage bins) can inform about the vehicles' mobility patterns and quality of the I2V service. In this paper, we show the contribution of POI (and site-specific information) to I2V transfers, leveraging a real-world dataset of geo-referenced I2V WiFi link measurements in urban settings. We present the distributions of throughput with respect to distance per POI class and site, and apply exponential regression to obtain practical throughput/distance models. We then use these models to compare I2V transfer estimation methodologies with different levels of POI-specific data and data resolution. We observe that I2V transfer estimate accuracy can improve from an average over-estimation of 18.3% with respect to measured values, if site or POI-specific information metrics are not used, to 9.3% in case such information is used.**

***Index Terms*—Vehicular networks, IoT nodes, I2V links, volume estimation**

I. Introduction

Smart Cities aim to improve the quality-of-life of city dwellers by increasing the efficiency of urban processes (e.g., traffic management), leveraging remote process monitoring and actuation by a centralized command center. The Internet-of-Things (IoT), a paradigm of pervasive and heterogeneous communication, is the key enabler of the data and control connections between sensors, actuators and control center. In this context, vehicular backhauls – i.e., vehicular fleets equipped with wireless Access Points (APs) – can also be explored to support dependable communication: data produced by road-side IoT nodes is opportunistically transferred to vehicles over infrastructure-to-vehicle (I2V) links and forwarded to the cloud [1]. During the design stage of an IoT platform, in order to select deployment sites for the road-side sensor/actuator nodes, it is necessary to estimate the I2V service offered by the vehicular backhaul at tentative deployment sites and evaluate it against the node's communication requirements. One option is to carry out measurement campaigns at all potential sites, but this can be resource-consuming or even unfeasible. Another option is to compute estimates of I2V transfers based on models of throughput versus distance and mobility traces, but these may be too demanding on computational resources and required datasets, or too simplistic thus failing to capture relevant behaviours.

In this paper, we explore how infrastructure features related to vehicular mobility can inform the estimation of I2V transfers. The characteristics of vehicular mobility (e.g., speed distribution and stopped/moving periods of the vehicles) can be related to local infrastructural points-of-interest (POI) that affect traffic (e.g., traffic lights, crosswalks) and/or fleet operation (bus stops, garbage bins). We leverage a dataset of geo-referenced I2V WiFi (IEEE 802.11b/g) link performance collected in three real-world urban settings to identify and quantify the relevance of POI. In an initial dataset analysis, we find that most connections occur when the vehicles are stopped. This lead us to apply a spatial clustering algorithm to the I2V samples, with the resulting clusters being associated with nearby POI. From this association, we draw distributions of throughput with respect to distance per class of POI (and target site), and apply regression (with an exponential curve) to produce a model that is more convenient for the estimation procedure. Finally, the resulting throughput/distance models are feed to a set of I2V transfer estimation methods that rely on POI and site-specific information to various degrees. In this manner, we identify the nature and scale of the datasets required for accurate estimation. We obtain an average over-estimation of 18.3% if non-specific models are not used, with improvements to 15% if POI-specific information is used and 9.3% if site-specific information is used.

Our main contributions are:

- Quantification of the contribution of vehicular point-of-interest to I2V transfers, in three representative real-world urban settings;
- Exponential models of throughput versus distance per type of POI and site obtained via regression;
- Comparative evaluation of I2V transfer estimation methods with different levels of coarseness and specificity.

The remainder of this article is organized as follows. In Section II, we review the literature on this topic. We describe the I2V dataset and pre-processing in Section III. In Section IV, we draw POI-related throughput and distance models at selected sites. Several estimation methods are evaluated in Section V. Final remarks are drawn in Section VI.

II. Related & Prior Work

One of the first reports of vehicular data collection from infrastructure nodes is found in [2]: the authors report a distributed mobile sensor system, in which data collected by sensors installed in vehicles is offloaded to static road-side units. The work of [3] builds on the previous by allowing vehicles to receive data from the static nodes. There are numerous works on IEEE 802.11 link characterization and data volume measurements in vehicle-to-X scenarios. The works reported in [4] and [5] evaluate the performance of IEEE 802.11-based I2V links to provide Internet to vehicular users in highway scenarios (in [5], also car-to-car measurements are taken). The authors of [6] describe range, association times, UDP and total data volume between infrastructure and cars at different speeds. In [7], the performance of IEEE 802.11 is characterized under aggressive mobility scenarios (node mobility up to 240km/h). Regarding I2V data volume estimation, the work of [8] presents a theoretical evaluation of the capacity and coverage of various technologies (cellular and vehicular) to support infrastructure-to-vehicle communication at large scale. The work of [9] discusses coverage estimation from cellular towers within the scope of Minimization Drive Tests (proposed by 3GPP), that seek to crowdsource user RSSI and position samples to support propagation estimation. The authors of [10] describe CARM, an algorithm to generate RSSI maps from crowdsensed datasets.

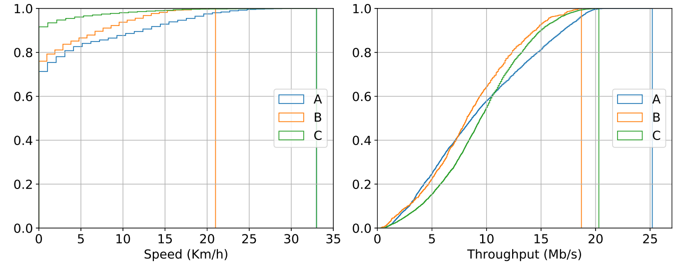
In previous work [11], we characterized transfer rates, volumes and connection times between a traffic light-bound sensor and passing buses over WiFi links. This work addressed a single site and the impact of nearby traffic and fleet-related elements was not studied in detail. In [12], we provide a first look into the relevance of POI over multiple sites: we introduce a dataset of geo-referenced I2V measurements with a waste disposal fleet and a methodology for sample clustering and POI association. In the current paper, we leverage the latter work (reviewed in Section III[1]) to obtain throughput models per site and POI, and compare various I2V volume estimation methods with different levels of requirements on input data detail and estimation accuracy.

III. Dataset Processing for Association with POI

A. Description of Geo-Referenced I2V Dataset

We use I2V measurements acquired in the context of the PortoLivingLab platform [13][2]. PortoLivingLab is a smart city-enabler IoT platform deployed in Porto, Portugal, comprising sub-platforms *UrbanSense* [14], a collection of 20 sensing units equipped with a WiFi module, and *BusNet*[3], a vehicular network of 600+ on-board units (OBU) installed in the public transportation and waste disposal fleets of Porto (the former accounting for 400 nodes) that offer a WiFi (IEEE 802.11b/g) hotspot service. The sensing units are configured as WiFi clients and connect opportunistically to passing OBUs, to perform I2V link performance measurement sessions. The measurements were taken at three sites refered to as **A**, **B** and **C**, and exclusively with the waste disposal fleet. The resulting logs contain collections of geo-referenced I2V measurement tuples (or simply I2V samples) composed of timestamp, MAC address of OBU, position and speed of vehicular node, and link

[1]Some values and figures have been updated per revised analysis processes.
[2]Dataset available on request to authors.
[3]BusNet is currently operated by company VENIAM: https://veniam.com/.



(a) Speed samples. (b) Throughput samples.

Fig. 1: CDF of relevant metrics for all sites.

quality metrics (throughput, packet loss ratio, jitter) collected with the tool *Iperf* (version 2.0). The time between I2V samples is at least 2s, as the vehicle's GPS position is obtained via a SSH-based query to the OBU, and the IPerf measurement session is scheduled to last 1s. Individual connections to OBU are identified by aggregating samples that are temporally close (less than 60 seconds apart). The data volume of a connection is the summation, over all samples, of the product of each sample throughput and the time interval until the next sample. In pre-processing, samples lacking valid GPS data were discarded. In total, we identified 12369 link quality samples, 588 connections, and 16 unique OBU.

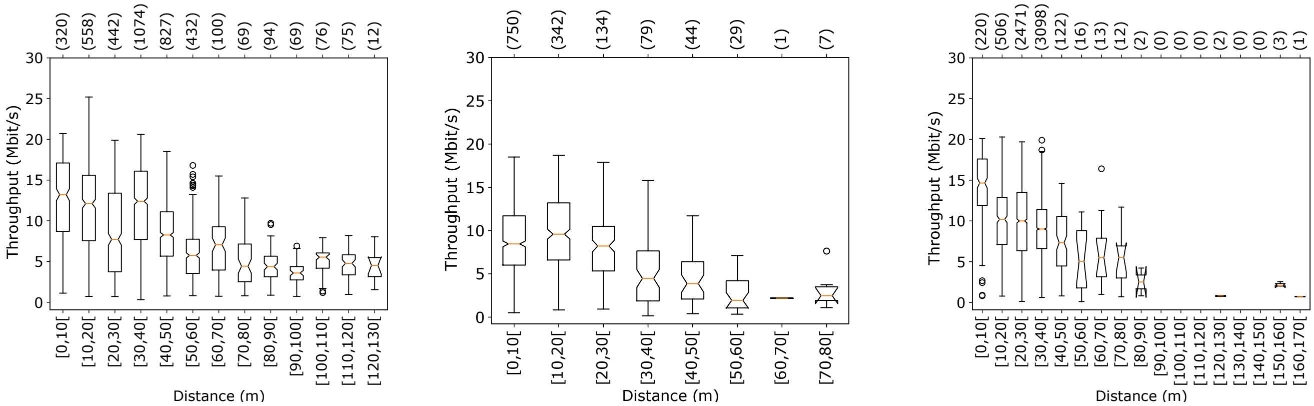
B. Stopped/Moving Ratio as Indicator of POI Impact

Using the geo-referenced I2V samples, we computed the per-site distributions of the vehicles' speed, throughput, and throughput per distance. The cumulative distribution function (CDF) of the speed of the vehicles during the I2V measurements, shown in Figure 1a (as in [12]), informs that the ratio of stopped and moving intervals of the vehicles was, on average, 83.32%. This leads us to conclude that the large majority of samples was taken when the vehicle is stopped, across all three sites. We also observed the throughput samples to follow similar distributions at the three sites, as can be seen from the throughput CDF in Figure 1b (as in [12]). The performance of throughput versus distance is depicted in Figure 2 for 10 meter-wide bins (as in [12]). We observe that the communication range differs between sites, which can be explained by distinct road topologies, and that there are distance intervals where a larger number of samples occur (see top axis of graphs), indicating that at those distances there may be points-of-interest (or areas affected by nearby POI).

Mobility dynamics can change due to circumstantial conditions (e.g., weather, emergency scenarios, holidays); in turn, these may cause the distribution of the contact intervals between the vehicle and the road-side wireless node to vary. Evaluating the impact of such conditions is challenging, as the involved steps are complex: (i) include external datasets to identify occurrences of such conditions; (ii) associate occurrences with changes in mobility patterns; and (iii) relate mobility pattern changes with I2V performance. Thus, we do not consider them in the current work.

C. Creation of I2V Clusters and Association to POI

In order to relate POI and spatial patterns of I2V samples, we create geographical clusters of samples that we subsequently map to the POI that could best motivate them. The



(a) Site A (b) Site B (c) Site C

Fig. 2: Throughput versus distance at the selected locations (top axis: number of samples).

Site	# days		Total	Stopped Filter	# clusters	Data retained in clusters*	Data retained in POI*	Assigned clusters
A	70	# samples	4148	3128 (75.4%)	30	35.9% (1123)	24.6% (768)	29
		Conn. time (s)	12825	9793 (76.4%)		31.5.0% (3084)	22.4% (2192)	
		Volume (Mbit)	103006	76614 (74.4%)		47.6% (36488)	29.2% (22364)	
B	34	# samples	1386	1124 (81.1%)	11	32.4% (364)	32.4% (364)	11
		Conn. time (s)	4094	3340 (81.6%)		36.0% (1202)	36.0% (1202)	
		Volume (Mbit)	31404	25995 (82.8%)		43.3% (11244)	43.3% (11244)	
C	61	# samples	6466	6043 (93.5%)	10	40.3% (2433)	38.1% (2302)	8
		Conn. time (s)	19875	18655 (93.9%)		32.7% (6102)	32.3% (6020)	
		Volume (Mbit)	170167	159813 (93.9%)		47.5% (75866.3)	46.9% (74877)	

**Ratios respective to the stopped dataset.*

TABLE I: Relevant metrics through data processing stages (stopped samples filtering, clustering and POI association).

first step of data processing was to keep only I2V measurement samples with stopped vehicles; we consider a vehicle as **stopped** if its speed is inferior to 3 km/h. The second step involved the application of a density-based clustering technique, DBSCAN [15], to identify areas of large sample density. In order to sub-sample homogeneously over the entire range of distances, we defined consecutive 10m-wide rings centered at the DCU location, and take only the 40% highest-throughput samples in each ring. We only apply this technique to the 40% highest quantiles of the throughput samples for two reasons: these are more indicative of locations favourable to I2V transfers, and throughput samples of low value can be found throughout the whole region with low density, thus hindering the clustering process. DBSCAN was then applied to the filtered I2V samples with the following parameters: neighborhood radius – 2.5m; minimum number of points to become a core point – 8. The parameters of the sub-sampling and DBSCAN procedures were found through empirical iterative exploration. Our target was to obtain a manageable number of clusters and with a size proportional to the surroundings (as opposed to having many small clusters or few large ones).

In a parallel process, we identified manually, at each site, the following classes of POI: **traffic lights** and **crosswalks** (traffic-related POI), and **garbage bins** (fleet-related POI). We then sought to assign clusters to POI, under the rule that one POI may be assigned more than one cluster (i.e., a POI can cause I2V sample clusters in different regions; the inverse is considerably more difficult to identify). We initially computed the Euclidean distance between cluster centroids and POI locations and associated them by closeness, but concluded that this approach did not perform well in some cases. E.g., oftentimes a slender cluster, known to be caused by a particular traffic light, presents its centroid closer to the crosswalk of an inflowing parallel street. Thus, we associated manually POI to clusters, considering: (a) whether POI is inside/near the cluster; (b) direction of traffic flows.

Table I (as in [12], updated) presents a quantitative characterization of the datasets associated with each processing step – the initial dataset, after the stopped filtering, after clustering, and after association to POI. The cluster algorithm retained between 32.4% and 40.3% of the stopped samples per site, corresponding to between 43.3% to 47.6% of the measured volume transfers in stopped conditions. This value range can further be altered by tuning the parameters of the sub-sampling and DBSCAN procedures. After association to POI, almost all samples are retained except at site A, where one large cluster could not be assigned to a POI and caused a considerable drop in the number of retained samples.

IV. Analysis and Models per POI Class and Site

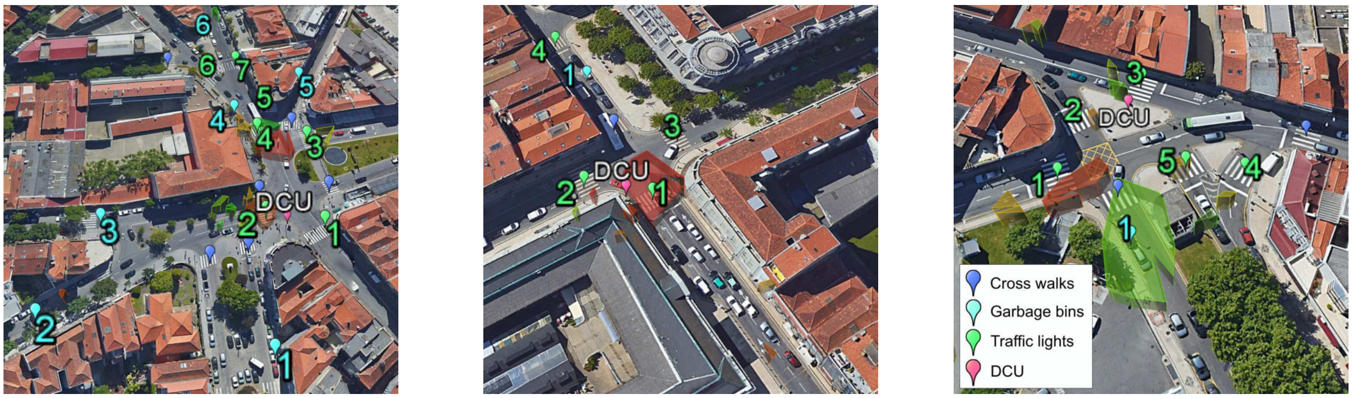
Using the association between clusters of I2V samples and POI described previously, the contribution of each POI class (i.e., the set of all POI of a given type [traffic-light, crosswalks, garbage bins]) at each site can be clarified.

A. I2V Service per Site and Relevance of POI Classes

We identify relationships between POI classes and I2V performance metrics that are site-specific. Table II (as in [12], updated) indicates, per site, the number of POI per class that were manually identified, the number of clusters associated to each POI class, and the corresponding dataset size and data volume. Figure 3 complements Table II by pinpointing traffic lights (green pins), garbage bins (blue pins), and produced clusters. Note that the *crosswalk* POI class refers only to crosswalks that are not related to traffic lights.

Site	POI class	# POI	# clusters	# samples	Conn. time at class POI (s)	Volume at class POI (Mbit)
A	Garbage bins	6	10	181	621	3719
	Traffic Lights	7	15	523	1409	16980
	Crosswalks	7	4	64	162	1666
B	Garbage bins	1	0	0	0	0
	Traffic lights	4	11	364	931	11244
	Crosswalks	1	0	0	0	0
C	Garbage bins	2	1	2064	5417	65598
	Traffic Lights	5	7	238	603	9378
	Crosswalks	2	0	0	0	0

TABLE II: Contribution of different POI classes for site performance.



(a) Site A (b) Site B (c) Site C

Fig. 3: Clusters of I2V samples (polygons) and POI (traffic lights and garbage bins are numbered).

Site A: most clusters are associated with traffic lights (e.g., note traffic lights [green pins] numbered as 2, 3, 4 and 5 of Figure 3a), accounting for 83% of all transfers. Two clusters are associated with garbage bins (2 and 6 [blue pins]). Crosswalks account for comparatively few samples. There is a large cluster that cannot be directly associated to any POI, as it sits at the center of the intersection.

Site B: the entirety of the data volume in this site is recorded at clusters associated with two traffic-lights (1 and 2 [green pins] in Fig. 3b). There is a garbage-bin POI, but no nearby cluster was produced during the clustering stage.

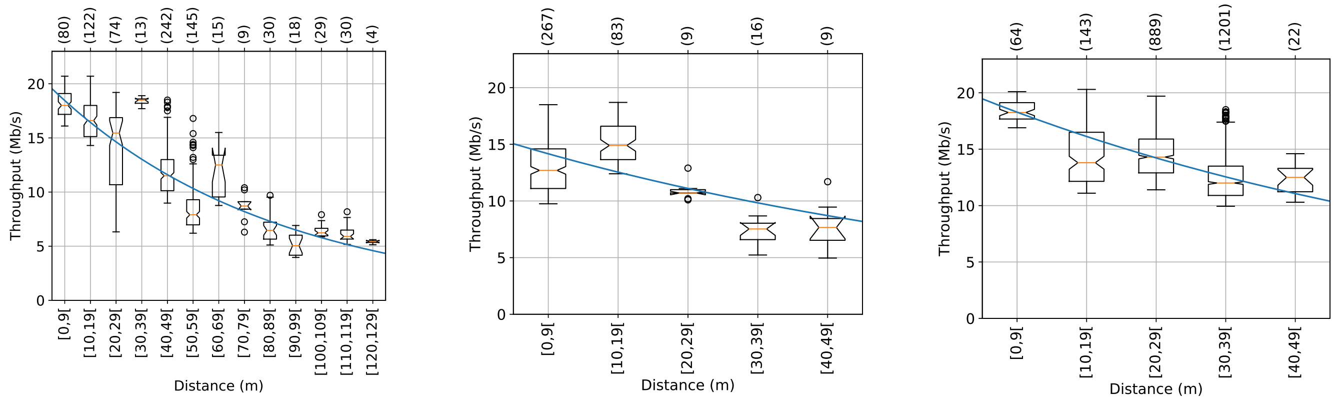
Site C: the majority of samples (87.5%) are associated with garbage bin 1 of Figure 3c. This is likely due to: (a) waste disposal trucks stop at this bin for long periods; (b) samples from nearby traffic lights might have been included.

In conclusion, we find that, in all sites, traffic lights always show associated samples and crosswalks account for few or no I2V samples. Also, the identification of a cluster at the center of the intersection, in site A, shows that other unaccounted factors may exist. Most samples at this cluster were obtained during night, thus excluding traffic jams as a cause. Finally, a single fleet-related POI may account of the majority of throughput samples at a given site (e.g., site C), to an extent that is not observed in traffic-related POI.

B. Throughput Models per POI Class and Site

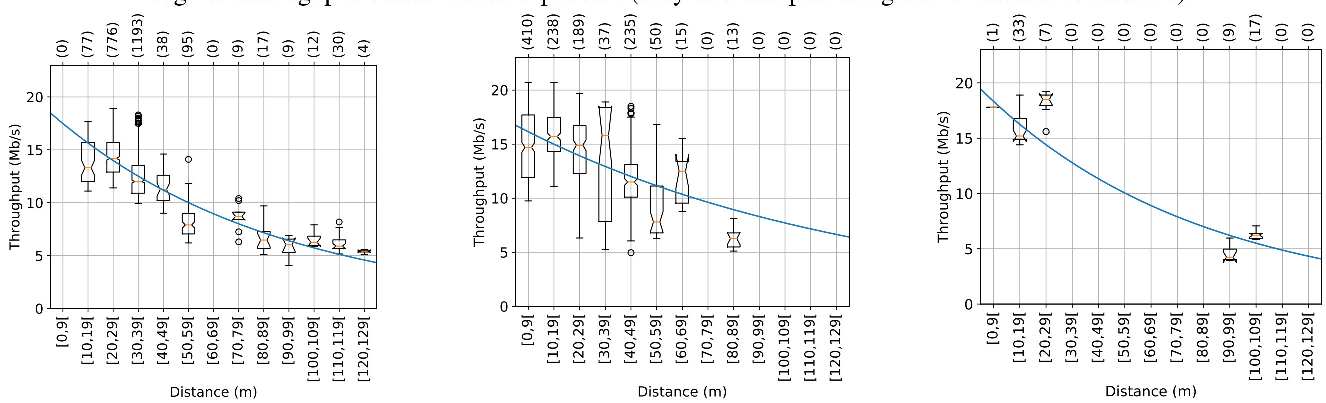
We now address the distribution of the throughput with respect to distance of the I2V clusters per site and POI class. The I2V samples assigned to clusters were binned in 10 meter-wide bins, shown in Figures 4 and 5 per site and POI class respectively (note that Fig. 4 differs from Fig. 2 as not all of the initial samples are assigned to clusters). In Figure 4, the gaps or distance ranges with inferior amount of samples further stress that each site presents its very own spatial configuration of POI. The results per POI class (Figure 5) capture some particularities of the behaviour of throughput with respect to distance at each type of POI, even if it is profoundly associated to the topology of the studied sites. For example, there are few samples collected at traffic lights that are far (i.e., above 90m), indicating that those traffic lights provide little contribution, while far away garbage bins still register considerable I2V exchanges. A challenge of POI class throughput characterization is that, even if more data is collected over more sites, it is unlikely (or very difficult) to obtain a dataset of I2V samples that are homogeneously distributed over the entire distance range.

In order to leverage the presented data for I2V volume estimation, we produce a concise and practical model of throughput with respect to distance ξ by fitting an exponential curve $t(d) = \alpha \cdot e^{\beta\ d}$ (t=throughput; d=distance) [16]. We computed model parameters for all data (ξ), for each site $\mathbf{S}$ ($\xi_{\mathbf{S}}$) – i.e., considering all I2V samples of each site –, for each POI class $\mathbf{P}_k$ ($\xi_{\mathbf{P}_k}$, where k is an index for POI classes) – i.e., considering the clusters associated with each POI class across all sites –, or both ($\xi_{\mathbf{S},\mathbf{P}_k}$). For all data, the global throughput model ξ has curve parameters α=17.5 and β=-0.0094; per-site and per-POI class models are shown in Table III. Looking at the site-related models $\xi_{\mathbf{S}}$ (Table IIIa), all sites present a similar decay factor $\beta \in [-1.26, -1.16] \cdot 10^{-2}$. Sites A and C exhibit similar throughput at 0m α (19.55 and 19.47 Mbit/s), whereas site B only attains 15.05, possibly indicating that communication at small distances is underperforming (e.g., signal may be attenuated by obstacles due to the position of the wireless interface). We also observe relevant differences among POI classes (Table IIIb, row $\xi_{\mathbf{P}_k}$). Garbage bins and crosswalks present the highest throughput at 0m α (18.49 Mbit/s and 19.47 Mbit/s, respectively), with respect to the inferior values of traffic lights (16.75 Mbit/s). The decay factor β presents values in the range ($\beta \in [-1.2, -0.74] \cdot 10^{-2}$), with traffic lights presenting the least steep decay, possibly due to the lack of samples at distances higher than 90m. The remaining



(a) Site A (b) Site B (c) Site C

Fig. 4: Throughput versus distance per site (only I2V samples assigned to clusters considered).



(a) Garbage bins (b) Traffic lights (c) Crosswalk

Fig. 5: Throughput versus distance per POI class (only I2V samples assigned to clusters considered).

(a) Per site.

Site		A		B		C	
Params.		α	β	α	β	α	β
$\xi_{\mathbf{S}}$		**19.55**	**-0.0116**	**15.05**	**-0.0122**	**19.47**	**-0.0126**

(b) Per POI class, global and at each site.

POI class		Garb. Bin		Traf. light		Crosswalk	
Params.		α	β	α	β	α	β
$\xi_{\mathbf{P}_k}$		**18.49**	**-0.0112**	**16.75**	**-0.0074**	**19.47**	**-0.012**
$\xi_{\mathbf{S},\mathbf{P}_k}$	Site A	11.26	-0.0053	19.79	-0.0112	19.47	-0.012
	Site B	-	-	15.05	-0.0122	-	-
	Site C	18.92	-0.0118	18.92	-0.0092	-	-

TABLE III: Exponential parameters for throughput models.

rows of Table IIIb present the models specific per site and POI class $\xi_{\mathbf{S},\mathbf{P}_k}$. We evaluate next if the per-POI class behaviours captured in these models can improve volume estimation.

V. I2V Estimation Methods

We compare several estimation methodologies with different levels of granularity and requirements on POI- or site-related detail, leveraging the developed I2V throughput models. As volume estimation may be applied in locations where datasets as the one of this work are not available, we seek to map levels of dataset detail and POI-/site-specific information into levels of estimation accuracy. In this way, parties interested in deploying road-side nodes that explore I2V transfers (e.g., municipalities, authorities, services), but that do not have the ability to produce large datasets of geo-referenced link performance samples, may learn about the accuracy that coarser and simpler-to-obtain datasets can offer. The following methods are formulated in a generic manner in order to be independent of specific data sources; in Subsection V-B, we clarify how our dataset was processed to implement them.

A. *Estimation Methodologies*

We seek to estimate the total data volume $\hat{V}$ at a tentative deployment location (referred to as the test location) in a given site $\mathbf{S}$. Four estimation models are presented next and assigned names, for convenience in the discussion. The index i identifies individual clusters, whereas k indexes a concrete POI class. The terms *POI* and *cluster* will be used interchangeably, but recall that a POI may have more that one cluster associated (i.e., a traffic light may cause stopping behaviours in distinct geographical areas). The used throughput models are those described in the previous section; Γ refers to total connection time. In all cases, it is assumed there is a prior geographical identification of stopping areas, that in turn correspond to potential clusters. This can be done e.g. by on-site inspection of existing POI and deduction of potential stopping areas caused by each POI, or mined from GPS traces if available.

Model 1, *Coarsest model* (Eq. 1): is agnostic to POI classes and builds on a basic set of information: an estimate of overall stopped time at the site $\underline{\Gamma}$, a global throughput model ξ, and a list of individual distances between cluster centroids and test location. The distance between cluster centroid/test location is used with the global throughput model to obtain a throughput value per cluster; this set of values is then averaged ($\overline{\xi}$).

$$\hat{V}_{\mathrm{M1}} = \underline{\Gamma}\ \overline{\xi} \tag{1}$$

To produce this estimate, an interested party could estimate $\underline{\Gamma}$ from traffic light scheduling; the throughput model ξ can be conveniently drawn from an existing representative collection of I2V measurements taken throughout the city, or a generic literature model; and the cluster centroid/test location distances can be drawn from a prior survey of potential clusters.

Model 2, *Site/POI model* (Eq. 2): uses a mix of POI class- and site-specific information: the total stopped time per POI class $\underline{\Gamma}_k$ and the average throughput for that site $\overline{\xi}_{\mathbf{S}}$. As in Model 1, the average throughput is obtained from feeding cluster centroid/test location distances into the site-specific throughput model and averaging the resulting values.

$$\hat{V}_{\mathrm{M2}} = \sum_k\ \underline{\Gamma}_k\ \overline{\xi}_{\mathbf{S}} \tag{2}$$

In practical terms, the total stopped time per POI class can be drawn from POI-related task execution time logs (e.g., in the case of waste disposal, average time at each garbage bin), or from GPS traces of fleets that use that particular class of POI. The average throughput per site $\overline{\xi}_{\mathbf{S}}$ relies on throughput measurements (not necessarily paired with position information) that can be done on purpose at a tentative deployment site.

Model 3, *POI-only model* (Eq. 3): considers metrics only related to POI classes, namely the total stopped time $\underline{\Gamma}_k$ and the average throughput for that POI class $\overline{\xi}_{\mathbf{P}_k}$.

$$\hat{V}_{\mathrm{M3}} = \sum_k\ \underline{\Gamma}_k\ \overline{\xi}_{\mathbf{P}_k} \tag{3}$$

This method is suited for the case in which fleets are wireless-enabled. The throughput model per POI class $\overline{\xi}_{\mathbf{P}_k}$ requires I2V measurements taken when the vehicle was stopped due to a POI of that class – again, this can be done by task execution timestamps, and not necessarily drawn from GPS data. $\underline{\Gamma}_k$ can be obtained as described in Model 2.

Model 4, *Per-cluster model* (Eq. 4): provides the finest resolution: the stopped time measured at each cluster Γ_i is individually multiplied by the site- and POI class-specific throughput $\xi_{\mathbf{S},\mathbf{P}_k}$ at the distance between cluster centroid and test location.

$$\hat{V}_{\mathrm{M4}} = \sum_k \left(\sum_{i\in\mathbf{P}_k} \Gamma_i\ \xi_{\mathbf{S},\mathbf{P}_k} \right) \tag{4}$$

This approach requires a way of estimating stop times at all clusters, to an extent that GPS data may be required. The site- and POI class-specific throughput may be obtained from on-site measurements at the cluster centroids (requiring already an extensive measurement campaign).

B. Evaluation of I2V Volume Estimation Models

We leverage our dataset to produce estimates according to the various models and compare them against the measured data volume in each site (recall that the described methods do not necessarily need a dataset like the one used in this article). We explain next how our dataset was processed to match the inputs of each method; the test locations are the locations of the DCUs where the measurements were taken.

Model 1: From our dataset, we compute the total stopped time of all clusters Γ_i in the target site, and multiply it by the average throughput predicted by the general throughput/distance model $\xi(d_i)$ for that site, thus producing a total volume estimate $\hat{V}'_{\mathrm{M1}}$.

$$\hat{V}'_{\mathrm{M1}} = \left(\sum_i \Gamma_i \right) \left(\frac{1}{N} \sum_i \xi(d_i) \right) \tag{5}$$

Model 2: We compute the product of the total stopped time per POI class ($\Gamma_i,\ i \in \mathbf{P}_k$), and the throughput estimated from the site-specific throughput/distance model $\xi_{\mathbf{S}}$ (at the distance between test location and the cluster center). The used exponential model parameters are those of Table IIIa. Summing the values per POI class, we produce the total volume estimate $\hat{V}'_{\mathrm{M2}}$.

$$\hat{V}'_{\mathrm{M2}} = \sum_k \left(\sum_{i\in\mathbf{P}_k} \Gamma_i \right) \left(\frac{1}{N} \sum_{i\in\mathbf{P}_k} \xi_{\mathbf{S}}(d_i) \right) \tag{6}$$

Model 3: Similar to Model 2, but the used throughput/distance model is defined per POI class, $\xi_{\mathbf{P}_k}$; the parameters are found in the respective line of Table IIIb.

Model 4: Eq. 4 was used directly with the model parameters of Table IIIb.

Table IV presents the data volume estimated by the models and the values measured on site, and the ratios to the volumes measured from samples in POI clusters and from all samples at each site. Regarding the ratios to POI cluster volumes (Table IVb, first three columns), Model 1 shows the highest over-estimation – between 1.11 and 1.25 (avg. 1.183) –, due to the very coarse temporal and throughput resolution used (in both cases, global metrics) and lack of site or POI class specific information. Model 2 exhibits an inferior over-estimation with respect to Model 1, between 1.03 and 1.16 with respect to measured values (avg. 1.093); the use of site-specific throughput metrics and POI class-aware time sums improved the estimation quality. Model 3, that incorporates POI class-related information, produces ratios between 1.08 and 1.23 (avg. 1.15), thus performing slightly worst that Model 2. Model 4, using the most specific models, shows mixed performance with an over-estimation over the actual values between 1.09 and 1.18 (avg. 1.127). Finally, the last three columns of Table IVb show that the estimates fall between 0.24 and 0.55 of the total volume of each site. We discuss these results in more detail in the following subsection.

C. Discussion

The end goal is to determine if POI can be predictors of the I2V transfers that a road-side wireless node installed in a given site will experience. We observe that coarse characterizations as those of Model 1 produce an over-estimation of up to 25% with respect to the volumes transferred in POI clusters. The inclusion of site-specific information, such as throughput-related metrics, improved estimation accuracy: Model 2 presents an over-estimation of up to 16%, with good performance on sites B and C (3% and 9% respectively). The use of POI class-specific average metrics (throughput and stop time), as in

(a) Measured and estimated volumes (Mbit).

Site (days)		A (70)	B (34)	C (61)
Measured	Total	103006	31404	170167
	POI clusters	22364	11244	74877
Estimated	*Coarsest model* (M1)	24889	14008	89310
	Site/POI model (M2)	25956	11562	81824
	POI-only model (M3)	25519	13840	80515
	Per-cluster model (M4)	26449	12536	81570

(b) Ratio between estimated and measured volumes.

Ratio	*To POI clusters volume*			*To total volume*		
Sites	A	B	C	A	B	C
Coarsest model (M1)	**1.11**	**1.25**	**1.19**	0.24	0.45	0.55
Site/POI model (M2)	**1.16**	**1.03**	**1.09**	0.25	0.37	0.48
POI-only model (M3)	**1.14**	**1.23**	**1.08**	0.25	0.44	0.47
Per-cluster model (M4)	**1.18**	**1.11**	**1.09**	0.26	0.4	0.48

TABLE IV: Model estimates compared to measured values.

Model 3, has mixed effects. In sites A and C there is an improvement in the estimate accuracy, but considerable degradation in site B. This may stem from the fact POI class-specific throughput-related metrics may not be as representative as site-specific metrics, as they are limited to the distances at which POI are found throughout the studied sites. The use of models tailored to site and POI class, as in Model 4, brings some improvement, but not substantial with respect to Model 2. This may be due to the lack of sufficient samples, causing the resulting models to not be accurate throughout the entire range of distances. As a practical takeaway, the observation that coarse metrics provide reasonable estimates indicates that less expensive techniques for measuring throughput and stopped time can be explored when producing datasets for I2V estimation.

Finally, the general behaviour of under-estimation of the total volumes (last three columns of Table IVb) is strongly related to the data retention rates of the clustering procedure. In designing the procedure, we aimed to find a trade-off between number of clusters and their geographical size, at the cost of some I2V samples not being assigned to clusters and thus not being considered by the estimation methods. This is a challenge bound to occur in any work that attempts to relate geo-referenced I2V samples and infrastructural elements. Thus, the ratios shown should not be used to evaluate the accuracy of the estimation methods, but they inform about the ability of the overall system in producing estimates of the total volumes that can be experienced at a potential site, under the current values of the clustering procedure parameters.

VI. Conclusions and Future Work

From analysis of real-world geo-referenced I2V link measurements with fleet vehicles, we observed that I2V transfers occur mostly with stopped vehicles. We created geographical clusters of I2V measurements and associated them with infrastructural point-of-interests relating with traffic or fleet operation. Exponential curves are fitted to the throughput samples with respect to distance per POI class and site. We observed relevant differences between the POI class-specific model parameters, e.g., traffic lights present a inferior throughput at small distances and a less steep decay factor. Finally, we show that the I2V volume estimation improves as site or POI-specific information is used: if no specific information is used, we report an average over-estimation of 18.3% with respect to measured values, whereas the incorporation of site- and POI-specific models allows to decrease it to 9.3% and 15%, respectively.

Future work will deepen the relationship between POI and measured I2V transfers at a given site. The setup of a second IoT node at one of the studied sites may clarify if our approach and conclusions are generalizable. A more comprehensive and generic model, capable of approximating the throughput versus distance behaviour at arbitrary locations, can be built by collecting additional measurements at other sites and under a wider variety of conditions. A clustering mechanism capable of retaining an high ratio of I2V samples without degrading the geographical boundaries of clusters will lead to better estimates. The emergence of new wireless technologies such as mmWave (IEEE 802.11ad) and multi-user MIMO (IEEE 802.11ax) may demand updated estimation mechanisms. Our long-term goal is to identify the minimum set of site characterization required to produce accurate estimates of I2V capacity.

VII. Acknowledgements

This work is a result of the Projects MobiWise (SAICTPAC/0011/2015, POCI-01-0145-FEDER-016426), S2MovingCity (CMUP-ERI/TIC/0010/2014), and the AQUAMON project (PTDC/CCICOM/30142/2017), as well as the CISTER (UID/CEC/04234) and IT (UID/EEA/50008/2019) Research Units, funded by the applicable financial framework: Fundo Europeu de Desenvolvimento Regional (FEDER), Programa Operacional Competitividade e Internacionalização (POCI), FCT/MCTES (Portuguese Foundation for Science and Technology).

References

[1] P. M. Santos, T. Calçada, D. Guimarães, T. Condeixa, S. Sargento, A. Aguiar, and J. Barros, "Demo: Platform for collecting data from urban sensors using vehicular networking," in *21st ACM MobiCom, Paris, France*. New York, NY, USA: ACM, 2015, pp. 167–169.

[2] B. Hull, V. Bychkovsky, Y. Zhang, K. Chen, M. Goraczko, A. K. Miu, E. Shih, H. Balakrishnan, and S. Madden, "CarTel: A Distributed Mobile Sensor Computing System," in *4th ACM SenSys*. New York, NY, USA: ACM, Nov. 2006, pp. 125–138.

[3] J. Eriksson, H. Balakrishnan, and S. Madden, "Cabernet: Vehicular content delivery using wifi," in *14th ACM MobiCom*. New York, NY, USA: ACM, 2008, pp. 199–210.

[4] J. Ott and D. Kutscher, "Drive-thru internet: Ieee 802.11b for "automobile" users," in *23rd Conference of IEEE Computer and Communications Societies, INFOCOM 2004*, vol. 1, March 2004, pp. –373.

[5] M. Wellens, B. Westphal, and P. Mahonen, "Performance evaluation of ieee 802.11-based wlans in vehicular scenarios," in *65th IEEE Vehicular Technology Conference*, 2007, pp. 1167–71.

[6] R. Gass, J. Scott, and C. Diot, "Measurements of in-motion 802.11 networking," in *7th IEEE Workshop on Mobile Computing Systems and Applications*, Aug 2006, pp. 69–74.

[7] P. Bergamo, M. Cesana, D. Maniezzo, G. Pau, K. Yao, D. Whiteman, and M. Gerla, "Ieee 802.11 wireless network under aggressive mobility scenario," in *Int'l Teletraffic Congress (ITC), Las Vegas, NV*, 2003.

[8] P. Belanovic, D. Valerio, A. Paier, T. Zemen, F. Ricciato, and C. F. Mecklenbrauker, "On wireless links for vehicle-to-infrastructure communications," *IEEE Trans. on Vehic. Tech.*, vol. 59, no. 1, pp. 269–282, Jan 2010.

[9] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, and B. Sayrac, "Spatial prediction under location uncertainty in cellular networks," *IEEE Trans. on Wireless Communications*, vol. 15, no. 11, pp. 7633–7643, Nov 2016.

[10] C. Xiang, P. Yang, C. Tian, L. Zhang, H. Lin, F. Xiao, M. Zhang, and Y. Liu, "Carm: Crowd-sensing accurate outdoor rss maps with error-prone smartphone measurements," *IEEE Trans. on Mobile Computing*, vol. 15, no. 11, pp. 2669–2681, Nov 2016.

[11] P. M. Santos, T. Calçada, S. Sargento, A. Aguiar, and J. Barros, "Experimental characterization of i2v wi-fi connections in an urban testbed," in *10th ACM MobiCom Workshop on Challenged Networks, Paris, France*. New York, NY, USA: ACM, 2015, pp. 5–8.

[12] L. M. Sousa, P. M. Santos, and A. Aguiar, "An exploratory study of relations between site features and i2v link performance," in *3rd EAI Int'l Conf. on IoT in Urban Space (Urb-IoT)*. Springer, Nov. 2018.

[13] P. M. Santos, J. G. P. Rodrigues, S. B. Cruz, T. Lourenço, P. M. dOrey, Y. Luis, C. Rocha, S. Sousa, S. Crisóstomo, C. Queirś, S. Sargento, A. Aguiar, and J. Barros, "Portolivinglab: An iot-based sensing platform for smart cities," *IEEE IoT Jrnl.*, vol. 5, no. 2, pp. 523–532, April 2018.

[14] Y. Luis, P. M. Santos, T. Lourenço, C. Pérez-Penichet, T. Calçada, and A. Aguiar, "Urbansense: An urban-scale sensing platform for the internet of things," in *2016 IEEE Int'l Smart Cities Conf. (ISC2)*, Sept. 2016.

[15] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *2nd Int'l Conf. on Know. Disc. and Data Min.* AAAI Press, 1996, pp. 226–231.

[16] L. R. Pinto, A. Moreira, L. Almeida, and A. Rowe, "Characterizing multihop aerial networks of cots multirotors," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 898–906, April 2017.