An Exploratory Study of Relations between Site Features and I2V Link Performance ⋆

Luís Sousa[1], Pedro M. Santos[1], and Ana Aguiar[1,2]

[1] Faculdade de Engenharia, Universidade do Porto
[2] Instituto de Telecomunicações
{lm.sousa,pmsantos,anaa}@fe.up.pt

Abstract. In emerging Internet-of-Things (IoT) realities, sensor nodes near roads will leverage opportunistic connections to vehicles to forward their data to the cloud. In planning such IoT platforms, node placement must be informed by an assessment of I2V transfers at a tentative location. It is not feasible to measure I2V volumes at all potential locations, so predictive models are necessary. We propose that qualitative characteristics of a potential site, in particular the existence of traffic and fleet-related points-of-interest (POI), can inform about the vehicles' mobility patterns and can ultimately be related with the quality of I2V service. In this work, we analyze a real-world dataset of WiFi I2V link measurements in an urban setting with an urban fleet. We observe that most connections occur with the vehicles stopped, and show that stopping regions are related with POIs. Our conclusions include that traffic lights and fleet POIs account for a considerable part of the collected I2V samples, whereas crosswalks account for few transfer occasions.

Keywords: Vehicular networks·IoT nodes·I2V links·volume estimation

1 Introduction

Smart Cities will require new solutions for connectivity and data collection inspired by the Internet-of-Things (IoT) paradigm [1]. Opportunistic collection of data produced by road-side IoT nodes by wireless-enabled vehicles is one of such strategies. When planning the deployment of such road-side nodes, an important challenge is to evaluate if the node will be able to transfer all of its data to passing vehicles. Learning this requires resource-consuming measurement campaigns at several potential locations, or accurate estimation methods of I2V transfers that rely on models of throughput vs. distance and mobility traces.

An alternative approach is to estimate data transfer rates and volume from relevant qualitative characteristics of a potential deployment location. Given the vehicular nature of the receivers, such features are typically related to mobility: these may be directly related to the vehicular mobility patterns (e.g., speed distribution and stopped/moving periods of the vehicles), or indirectly by mobility-affecting points-of-interest (POI), such as traffic (e.g., traffic lights, crosswalks) and/or fleet-related infrastructure (e.g., bus stops, garbage bins). In this paper, we identify mobility and throughput patterns that can be associated with such features, leveraging a dataset of WiFi link characterization in I2V connections. We observe that a large percentage of connection samples take place when the vehicles are stopped, and relate stopping areas to nearby POI.

Our main contributions are: (a) identification of mobility and throughput patterns associated to I2V service from analysis of a real-world link quality dataset; and (b) elation of patterns and POI of a site and break down per POI class.

⋆ This work is funded by project S2MovingCity, reference CMUP-ERI/TIC/0010/2014, funded by FCT/MEC (PIDDAC) through national funds, and IT internal funds, reference UID/EEA/50008/2013.

The remainder of this article is organised as follows. In Section 2 we review the literature on this topic. We describe the I2V dataset and experimental setup, and uncovered mobility and throughput patterns in Section 3. In Section 4, we relate identified mobility and throughput patterns with POIs of each site. Final remarks are drawn in Section 5.

2 Related Work

In [5], the authors report a distributed mobile sensor system: data collected by sensors installed in vehicles is offloaded to static road-side units. The work of [2] builds on the previous by allowing vehicles to receive data from the static nodes. The performance of I2V links based on IEEE 802.11 networks to provide Internet to vehicular users in highway scenarios has been characterized in [8]. The authors of [4] describe range, association times, UDP and total data volume between infrastructure and cars at different speeds. The tools to perform bandwidth estimation include the well-established Iperf application, that attempts to send the maximum data possible through the channel to estimate the channel capacity. WBest [6] offers an alternative for end-to-end capacity measurements: it assumes that end-to-end capacity is defined by the narrowest link, that "the time dispersion between the two packets is linearly related to that narrow link capacity" [6]. To the best of our knowledge, this is the first work proposing the estimation of I2V data transfers from site features.

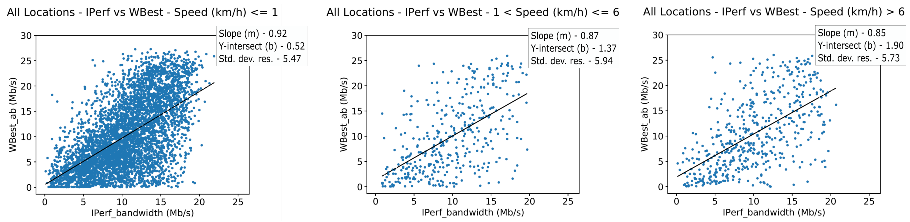
3 Experimental Setup and Dataset Analysis

We use I2V measurements acquired in the context of the PortoLivingLab platform [10]. PortoLivingLab is a smart city-enabler IoT platform deployed in Porto, Portugal, comprised by sub-platforms *UrbanSense* [7], a collection of 20 sensing units equipped with a WiFi module and named Data Collection Units (DCUs), and *BusNet*, a vehicular network of 600+ on-board units (OBUs) installed in the public transportation fleet (accounting for 400 nodes) and waste disposal fleet of Porto, and that offer a WiFi hotspot service. An initial characterization of I2V WiFi links in the PortoLivingLab setup is reported in [9].

The I2V experimental dataset was collected at three DCUs, deployed at three disparate sites and refered to as "A", "B" and "C", from measurement sessions carried out with the OBUs installed in the waste-disposal fleet. The DCUs, configured as WiFi clients, connected opportunistically to passing OBUs (configured as access points – AP) to collect the GPS of the vehicles and perform measurement sessions of Iperf and Wbest. The resulting dataset contains collections of samples (i.e., measurement tuples) composed of timestamp, MAC address of OBU, position and speed of vehicular node, and link quality metrics – throughput, PLR, jitter – from both tools. The three steps – collecting GPS and running Iperf and WBest measurement sessions – were performed sequentially and repeated in this order while the connection lasted. The time between tuples is at least 2s, as the GPS data of the vehicle is obtained via a SSH-based query to the OBU, and the IPerf measurement session is scheduled to last 1s. Individual connections to OBUs 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 obtained 12369 link quality samples, 588 connections, and detected 16 different OBUs.

3.1 Evaluation of Bandwidth Estimation tools

As mentioned earlier, the available dataset contains channel capacity measurements from two tools that operate in significantly different ways. Iperf injects the maximum possible data in the channel, and it can be highly disruptive to other on-going connections. WBest proposes a non-disruptive alternative, and thus it may be better suited when unrelated connections exist.



(a) Slow moving vehicles. (b) Medium speed. (c) Fast moving vehicles.

Fig. 1: Comparison of Iperf/WBest measured throughput, for the same measurement tuple.

We seek to understand how similar the estimates of both tools are. In the I2V dataset, the samples for each tool were obtained sequentially: the measurement session by Iperf, lasting 1s, is followed by a session by WBest. In Figure 1, we plot the relation of the Iperf and WBest values for the same measurement tuple, for the cases of stopped vehicles (speeds under 1 km/h), slowly moving (speed between 1 and 6km/h), and fast speed (over 6 km/h). Linear regression parameters and standard deviation of residuals are also shown. Observed slopes are close to 1 for most cases, particularly for slow moving vehicles. The slope decreases as the speed increases, indicating that WBest under-estimates the bandwidth, which in turn may be a consequence of increased packet loss. For the remainder of this paper, we use the Iperf measurements.

3.2 Analysis of Vehicle Mobility and I2V Throughput

We analyzed the collected data by evaluating how the measured link throughput varies with several mobility-related features, particularly the vehicles' speed and distance to the DCU. The speed of the vehicles during the I2V measurements is shown in Fig. 2. We observe that, across the three sites, the large majority of samples was taken when the vehicle is stopped. The ratio of stopped and moving intervals of the vehicles was, on average, 83.32%, so there are more throughput samples collected when the vehicles are stopped.

We also observed the throughput samples to follow similar distributions at the three sites, as can be seen from the throughput CDFs in Fig. 3. The performance of throughput versus distance is depicted in Fig. 4 for 10 meter-wide bins. We observe that the communication range differs from location to location, that can explained by the different 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 POIs (or areas affected by nearby POIs).

The two main takeaways are that: (a) in all locations, most of the connection samples occur when vehicles are stopped; (b) sample frequency w.r.t. distance varies between distance intervals and from location to location. These two observations lead us to conclude that, at each location, points-of-interest at particular distances create stopping opportunities.

4 Identification and I2V Performance at POI

We hypothesize, from the previous analysis, that most data transfers occur at locations where points-of-interest exist. We define two classes of POIs: (a) traffic-related, e.g., traffic lights, crosswalks, etc.; and (b) fleet-related, e.g., bus stop, garbage bins.

4.1 Identifying Stopping Regions and Association to POI

We further analyzed the I2V measurement samples at which vehicles were stopped. We consider a vehicle is "stopped" if its speed is inferior to 3 km/h. We verified by visual inspection that these samples were not uniformly distributed over the roadways, and that their spatial density is higher near POIs. We set out to confirm the link between POIs and "stopped" samples as follows.

In a first stage, we applied a density-based clustering technique, DBSCAN [3], to identify areas with high concentration of samples. The DBSCAN parameters were: neighborhood

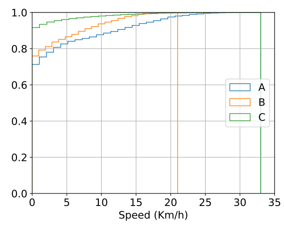


Fig. 2: CDF of speed samples for all sites.

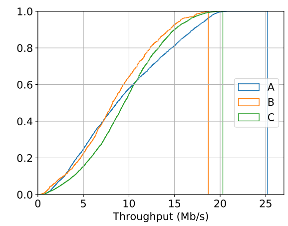
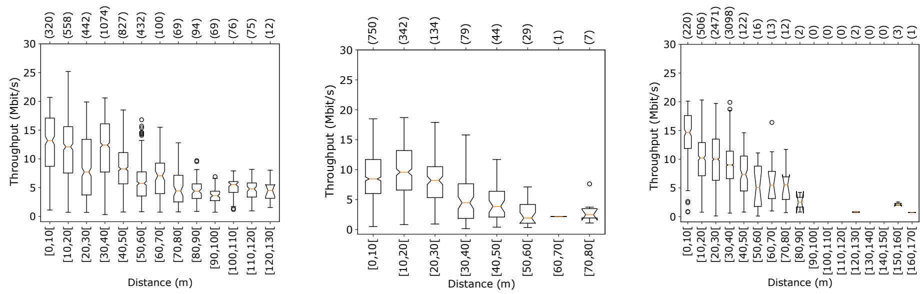


Fig. 3: CDF of throughput samples for all sites.



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

Fig. 4: Throughput vs. distance at the selected locations (number of samples on top axis).

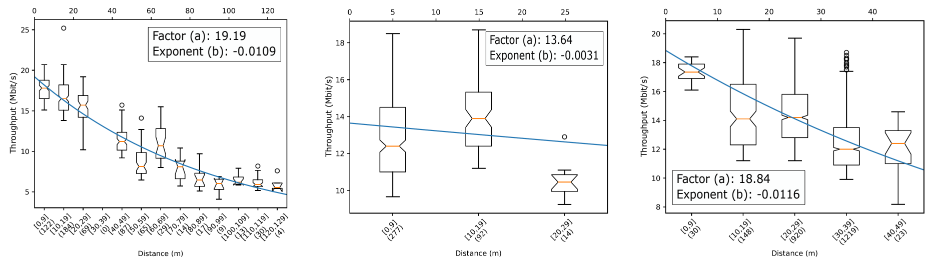
radius – 2.5m; minimum number of points to become a core point – 8. The original dataset was sub-sampled prior to the cluster algorithm, to retain the highest-throughput samples. We defined consecutive 10m-wide rings centered at the DCU location, and take only the 40% highest-throughput samples in each ring. This approach allows us to sub-sample homogeneously over the entire range of distances. The parameters of the sub-sampling and DBSCAN procedures were found after iterative search: at each round, we evaluated empirically the number and size of produced clusters. The goal was to obtain a manageable number of clusters for the subsequent POI assignment task (as opposed to having many small clusters or few large ones).

After the clustering algorithm was applied, we established an univocal correspondence between clusters and POIs. We identified manually, on each site, the following types of POIs: traffic lights, crosswalks and garbage bins. We computed the Euclidean distance between cluster centroids and POI locations, and associated the closest ones. We 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 a inflowing parallel street. Thus, we associated manually POIs to clusters, taking into consideration: (a) whether POI is inside the cluster; (b) direction of traffic flows. POIs may be assigned more than one cluster. Table 1 presents a quantitative characterization of the datasets corresponding to each of these steps – the initial dataset, and those associated with the produced clusters and assigned clusters. The cluster algorithm retained between 31.9% and 38.1% of the global dataset, corresponding to between 43.4% to 47.1% of volume transfers. Note that the clustering algorithm is only applied to stopped “samples”, and that the sub-sampling and DBSCAN procedures can be tuned to different results. After association to POIs, almost all cluster samples are retained, with the exception of site A where one large cluster could not be assigned to a POI (more details on the next section).

For the samples associated to POIs, we observe a consistent behaviour of throughput versus distance at most sites. Fig. 5 presents per-distance boxplots of transfer rates and a fitted exponential curve $t = a e^{bd}$, where t is throughput and d distance. Sites A and C, containing samples at least up to 50m, exhibit similar curves. Site B does not allow for extensive comparison (due to inferior distance range), but rates at the available distances verify those of other sites.

Site	# days		Total	Stopped	# clus.	Clusters	POIs	Assigned clusters
A	70	# samples	4148	75.4% (3128)	21	34.6% (1434)	16.5% (685)	18
		Conn. time (s)	12825	76.4% (9793)		31.1% (3984)	14.7% (1880)	
		Volume (Mbit)	103006	74.4% (76614)		45.2% (46528)	20.0% (20607)	
B	34	# samples	1386	81.1% (1124)	11	31.9% (442)	31.9% (442)	11
		Conn. time (s)	4094	81.6% (3340)		29.4% (1202)	29.4% (1202)	
		Volume (Mbit)	31404	82.8% (25995)		43.4% (13631)	43.4% (13631)	
C	61	# samples	6466	93.5% (6043)	10	38.1% (2466)	37.8% (2444)	8
		Conn. time (s)	19875	93.9% (18655)		32.4% (6444)	32.1% (6387)	
		Volume (Mbit)	170167	90.9% (159813)		47.1% (80109)	46.6% (79283)	

Table 1: Dataset metrics: total, stopped and cluster-bound (produced and assigned).



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

Fig. 5: Throughput vs. distance per POI class.

4.2 Analysis per POI Class and Location

We evaluate the contribution of each POI class to the data transfers measured at each site. Table 2 indicates the number of POIs per site, the number of clusters associated to each POI class, and the corresponding dataset size and data volume. Fig. 6 depicts the produced clusters and relevant POIs, for a visual reference to support the values of Table 2. Site-specific analyses are now drawn.

Site A: most samples and clusters are associated with traffic lights (e.g., 1, 3, 4 and 5 of Fig. 6a), and two clusters are associated with garbage bins (1 and 6). Despite being more numerous, crosswalks account for comparatively few samples. There is a large cluster that cannot be directly associated to any POI, as it is located at the center of the cross-roads.

Site B: the majority of the data volume in this site is recorded at clusters associated with two traffic-light (1 and 2 in Fig. 6b). There is a garbage bin POI, but no nearby cluster had been produced during the clustering stage.

Site C: the majority of samples are associated with garbage bin 1 of Fig. 6c. Trucks stop at this bin for long periods; samples from nearby traffic light might have been incorporated.

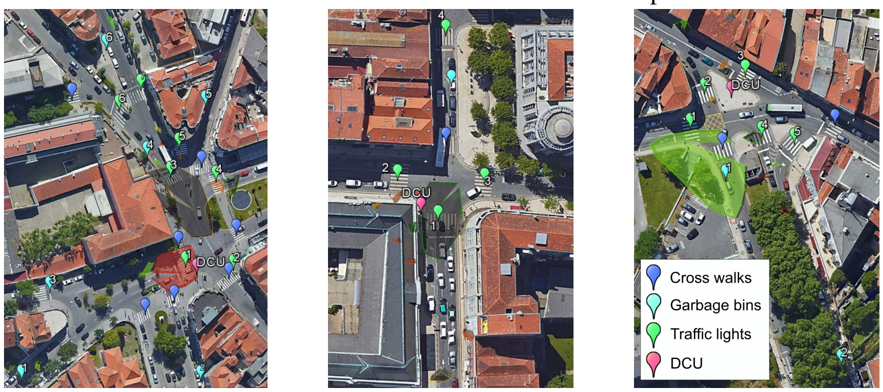
In summary, we observe a clear relation between POIs and the areas where most throughput samples of high value are recorded. Traffic lights always show associated samples, and cross-walks account for few or no I2V samples. The identification of a cluster at the center of the cross-roads, in site A, shows that other unaccounted factors may exist. Most samples at this cluster were obtained during night, thus excluding traffic jams as a cause; additional work is required to justify it. 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 POIs.

5 Conclusions and Future Work

Using real-world I2V link measurements, we observe that most measurements with fleet vehicles occur with stopped vehicles, and areas of relevant throughput samples (many samples of high value) are close to POI that affect the vehicles' mobility. In detail, the contribution to I2V transfers of traffic lights is always present, and that of crosswalks is residual; strategically-located fleet-related POIs can contribute considerably. The end goal is to build a estimation model of I2V data volumes based on qualitative features of a potential site.

Site	POI class	# POIs	# clusters	# samples	Conn. time at class POI (s)	Volume at class POI (Mbit)
A	Traffic Lights	7	9	548	1412	17023
	Crosswalks	7	2	60	152	1540
	Garbage bins	6	7	118	316	2044
B	Traffic lights	4	11	464	1202	13631
	Crosswalks	1	0	0	0	0
	Garbage bins	1	0	0	0	0
C	Traffic Lights	5	7	30	377	5424
	Crosswalks	2	0	0	0	0
	Garbage bins	2	1	2310	6010	73858

Table 2: Contribution of different POI classes for site performance.



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

Fig. 6: Clusters of I2V samples and POIs (traffic lights and bins are numbered).

References

1. Enterprise, H.: White paper: Smart cities and the internet of things. Tech. rep., HP (2016), http://h20195.www2.hpe.com/v2/getpdf.aspx/4AA6-5129ENW.pdf (last visited: 2017-01-11)
2. Eriksson, J., Balakrishnan, H., Madden, S.: Cabernet: Vehicular content delivery using wifi. In: Proceedings of the 14th ACM International Conference on Mobile Computing and Networking. pp. 199–210. MobiCom '08, ACM, New York, NY, USA (2008)
3. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. pp. 226–231. KDD'96, AAAI Press (1996)
4. Gass, R., Scott, J., Diot, C.: Measurements of in-motion 802.11 networking. In: 7th IEEE Workshop on Mobile Computing Systems and Applications. pp. 69–74 (Aug 2006)
5. Hull, B., Bychkovsky, V., Zhang, Y., Chen, K., Goraczko, M., Miu, A.K., Shih, E., Balakrishnan, H., Madden, S.: CarTel: A Distributed Mobile Sensor Computing System. In: 4th ACM SenSys. Boulder, CO (November 2006)
6. Li, M., Claypool, M., Kinicki, R.: Wbest: A bandwidth estimation tool for ieee 802.11 wireless networks. In: 33rd IEEE Conference on Local Computer Networks. pp. 374–381 (Oct 2008)
7. Luis, Y., Santos, P.M., Lourenco, T., Pérez-Penichet, C., Calçada, T., Aguiar, A.: Urbansense: An urban-scale sensing platform for the internet of things. In: 2016 IEEE International Smart Cities Conference (ISC2). pp. 1–6 (Sept 2016)
8. Ott, J., Kutscher, D.: Drive-thru internet: Ieee 802.11b for "automobile" users. In: 23rd Conference of IEEE Computer and Communications Societies, INFOCOM 2004. vol. 1, pp. –373 (March 2004)
9. Santos, P.M., Calçada, T., Sargento, S., Aguiar, A., Barros, J.: Experimental characterization of i2v wi-fi connections in an urban testbed. In: Proceedings of the 10th ACM MobiCom Workshop on Challenged Networks. pp. 5–8. ACM, New York, NY, USA (2015)
10. Santos, P.M., Rodrigues, J.G.P., Cruz, S.B., Loureno, T., dOrey, P.M., Luis, Y., Rocha, C., Sousa, S., Crisóstomo, S., Queirós, C., Sargento, S., Aguiar, A., Barros, J.: Portolivinglab: an iot-based sensing platform for smart cities. IEEE Internet of Things Journal **PP**(99), 1–1 (2018)