Opportunistic communication in smart city: experimental insight with small-scale taxi fleets as data carriers

Marco Bonola[a], Lorenzo Bracciale[a,*], Pierpaolo Loreti[a], Raul Amici[a], Antonello Rabuffi[a], Giuseppe Bianchi[a]

[a]*Electronic Engineering Department, University of Rome "Tor Vergata", Rome, Italy*

Abstract

How to deliver data to, or collect data from the hundreds of thousands of sensors and actuators integrated in "things" spread across virtually every smart citys streets (garbage cans, storm drains, advertising panels, etc)? The answer to the question is neither straightforward nor unique, given the scale of the issue, the lack of a single administrative entity for such tiny devices (arguably run by a multiplicity of distinct and independent service providers), and the cost and power concerns that their direct connectivity to the cellular network might pose. This paper posits that one possible alternative consists in connecting such devices to their data collection gateways using "oblivious data mules", namely transport fleets such as taxi cabs which (unlike most data mules considered in past work) have no relation whatsoever with the smart city service providers, nor are required to follow any pre-established or optimized path, nor are willing to share their LTE connectivity. We experimentally evaluate data collection and delivery performance using real world traces gathered over a six month period in the city of Rome. Results suggest that even relatively small fleets, such as an average of about 120 vehicles, operating in parallel in a very large and irregular city such as Rome, can achieve an 80% coverage of the downtown area in less than 24 hours.

Keywords: smart city, data mules, delay tolerant network, vanet, sensor network

*Corresponding author

Email addresses: `marco.bonola@uniroma2.it` (Marco Bonola), `lorenzo.bracciale@uniroma2.it` (Lorenzo Bracciale),

1. Introduction

Due to their extremely low costs, sensors and actuators can surely find their places in a wide range of applications in the Smart City. We can easily imagine, for instance, intelligent garbage cans that communicate their trash levels to a central operative base, or notifications sent by storm drains when they are clogged by leaves. These are only some of the multitude of examples in which small sensors can be plugged into existing objects to enhance their functionality or to improve productivity in general. However, this scenario presents the following two main challenges: i) how these sensors are powered; ii) how they can communicate with third entities for configuration and data storing.

Recent studies [1] reveal that energy harvesting is now more than an academic utopia. For instance, drain systems could acquire energy from people moving above them, while nodes installed on garbage cans could acquire energy from people throwing their garbage away. Depending on the specific case, sensor nodes can obtain energy in different manners (solar, piezoelectric, etc) that is, in many cases, more than sufficient for sporadic data sensing. *But is it also enough for communication?*

The amount of energy required for that issue could be critical although highly dependent on the quality of the communication we need. Real-time communication could indeed be very expensive because of the need of a permanent infrastructure to be built to cover all the nodes. If, on the one hand, interconnecting a fine-grained sensor network to the public internet or to other large scale IP network infrastructures may seem trivial from a technological point of view, on the other hand, it is undoubtedly difficult from both economic and administrative points of view to either equip each sensor device with a SIM card or disseminate over the entire city sensor gateways or femtocells managed by the same administrative entity. A real "smart" city should therefore be able to exploit and repurpose existing resources in order to minimize the infrastructure and management impact. Moreover, if these sensors are extremely energy constrained, their communication range can not be very large, hence complete coverage would also be technologically unfeasible.

Data Mules

The problem is all but new. An interesting way to address this issue is to introduce mobile nodes that act as "data mules" [2] and that completely cover

pierpaolo.loreti@uniroma2.it (Pierpaolo Loreti), raul.amici@gmail.com (Raul Amici), antonello.rabuffi@gmail.com (Antonello Rabuffi), giuseppe.bianch@uniroma2.it (Giuseppe Bianchi)

the interested area by passing near the deployed nodes and performing actions such as data collection or delivery. In fact, this solution *provides a mobile data infrastructure as a substitute for the fixed infrastructure*, suitable for many delay tolerant services [3] [4] [5] [6]. From an operating and administrative point of view, if we have several different services (e.g. waste, street lighting etc.) we need different management entities that control the mules, and assure that the specific quality of service is met. Otherwise, a single management entity (e.g. the municipality) could control and manage the mules on behalf of several service operators, taking advantage of the statistical multiplexing of the vehicles that moves in the city.
What happens if we do not want to place any controls on the mules?

Oblivious Data Mules

In this paper we do not want to use any type of managed infrastructure, neither fixed nor mobile, but rather rely on pure opportunistic connections between vehicles and nodes. In this regard, the huge number of public, semi-public, or private vehicles that circulate undoubtedly represents a unique opportunity for Smart Cities: these nodes that move in the city every day could act as IoT mobile nodes. Obviously, if all the vehicles in a city were involved, the quality of the communication service offered would be optimal. Reaching a high consensus in such an initiative is by no means easy in the short term. A more realistic alternative is to consider that today there are several public services that involve moving vehicles (buses, ambulances, mail delivery) - among these, the taxi service is probably one of the most interesting because cabs cover an extensive area of the city with somewhat random paths (with respect to buses that move only on main streets) and work 24/7. As a matter of fact, they also provide a better coverage of the part of the city that is more densely populated and that usually corresponds to the place where we need to collect and send more data.

Moreover, many taxi fleets have started using external ICT services for ride dispatching since they already have almost all the needed hardware and equipment on board; they are ideal candidates as **oblivious data mules**.

It is important to stress that the mules are truly "oblivious" since we do not want to force the cars to follow planned paths, nor to vary driving habits, and neither to share their internet connection if they do not want to. The mules have to be equipped with the technology needed to communicate with the devices of the IoT, a storage capacity, and a commodity internet connection (optional).

Hence, in the scenario we propose, these cars move independently from the position of sensor nodes but according to customers' demands, ready to get/push data but rather incidentally and opportunistically collect and distribute the data.

We can thus provide a temporary and fine-grained coverage of the territory without big expenses.

Contributions

In this paper we evaluate the feasibility and the performance of a scenario where we have **no guarantees** on and **no requisites** for the data communication service.

However, we will show how we can provide some statistical guarantee that will be good enough to enhance many specific services, given that the communication is provided almost "for free", boosting the IoT with new capacity.

In detail, the main contributions of this paper are:

- Assessing the technological feasibility and system performance of an infrastructure for sensor data gathering/dissemination composed of a relatively small taxi fleet used as data mules: we base our analyses on a real experiment that has been conducted in Rome, Italy, that represents the longest data collection available (6 months) with the finest sample rate (7 sec). To the best of our knowledge, this work is the first ever to validate data mule applications with real world mobility traces.

- Identifying the performance metrics related to the proposed service scenario, analysing the range of validity of the analysis with the aim of assessing the statistical guarantees that we can provide for this service scenario and how they vary in time.

- Evaluating different service scenarios for data dissemination and gathering from the data mules to the IoT nodes and vice-versa.

- Analysing some energy optimization techniques to show how the performance metrics are affected by a windowed and adaptive duty cycle.

In the remainder of this paper we will show that we can provide statistical guarantees to enhance many specific services, given that communication is provided almost "for free", boosting the IoT with new capacity. To the best of our knowledge, this work is the first to ever validate data mule applications with real world mobility traces.

2. Application scenario

2.1. Data Mules

We imagine a scenario where a small fleet of taxi cabs moves inside a city doing their job: waiting for rides in the car parks, picking up and shuttling customers. The taxi cabs are equipped with: an on-board terminal that integrates different wireless technologies to communicate the surrounding environment (WiFi, ZigBee, Bluetooth, RFID, etc.); a buffering capacity; and an optional internet connection. These terminals could be ordinary tablets or smartphones, equipped with a radio interface suitable for low power communication [1]

Moreover, we suppose that special areas in the city (for taxi cabs the most natural choice is to use their parking lots) can act as hotspots that provide internet access in case we do not have or we do not want to use the on-board internet connection. In this case we can move the data via the *store-carry-forward* routing technique.

Figure 1: Proposed scenario: taxi cabs incidentally exchange data with the nodes when they pass nearby.

The taxi cabs are our "data mules" that obliviously become part of the IoT opportunistically exchanging data with the sensor/actuator nodes deployed all over the smart city. As depicted in figure 1, while the taxi cabs move around in the city according to customers' requests and typical roaming routines, they incidentally cross the coverage area of active nodes with which they can exchange small quantities of data. Once the data is acquired from a node, it can be instantaneously sent to a main server via the on-board internet connection, or buffered

[1]Bluetooth Low Energy is already supported as a built-in functionality in Android 4.3 and on IOS 6 devices. IEEE 802.15.4 radio interface could be provided via sim cards [7] or with external adaptors.

and then delivered via hotspots, realizing a *data gathering* service. In turn, the on-board terminals could receive data from the internet or the hotspots and then diffuse that data to all the nodes the vehicle encounters, realizing a *data diffusion* service.

2.2. Pervasive nodes and services

We imagine that these sensors/actuators (hereafter called smart city nodes, or just nodes for brevity) are not internally powered and should acquire their energy from external sources via photovoltaic, piezoelectric or other energy harvesting solutions. Consequently, nodes are severely energy constrained. They can perform raw sensing and can sometimes communicate with their wireless technology; yet they can also suddenly run out of power.

The offered communication service is therefore *definitely best effort*, but has the great advantage that nodes can be very inexpensive and do not imply any extra fees (e.g. monthly SIM card costs or battery recharging/substitution), hence they can be plugged on top of other existing solutions to enhance existing services, even if only partially.

Examples of these services are:

- **Smart waste collection**: each garbage bin communicates its filling level to a central system that can optimize the collection routes and times
- **Street lights**: lights can be programmed to vary their on-off cycling according specific policies (e.g. daylight, fog, special events)
- **Storm drains**: storm drains can communicate their operating status to their maintainers, to signal overflow or obstruction (e.g. by leaves)

and many more. If we look at the communication needs of these services, they need small amounts of data every once in a while, also also tolerating significant delays. This is a very important issue that we exploit in proposing a different model of delay tolerant, opportunistic communications that do not require any fixed communication or power infrastructure, representing one of the prominent factors that today really prevent these services from being installed in every city. It is worth noting that in this work we do not consider unicast data communication, event if this is theoretically supported by the proposed application scenario.

3. Evaluation methodology

We evaluated the performance of the system using real traces, collected by periodically logging the time and positions of 320 cars of a taxi fleet driving in the city of Rome, Italy, for a period of 6 months. A one month window of these traces are available at the public wireless data archive Crawdad [8].

Scenario

We limit our analysis to the center of Rome, where the density of the taxi cabs is relevant.
Taxi drivers work 24/7 on shifts so that on average 120 drivers are working at the same time.
We consider an area of 8km x 8km whose bounds are given by the coordinate pairs (41.856, 12.442) (41.928, 12.5387).
This scenario is characterized by very narrow and congested roads, high traffic volume, and slow speed, as usually happens in city centers.
For the sake of simplicity, the area has been analysed using a 200x200 grid where a single grid cell covers a square area of 40x40m. We assume that when a taxi cab enters a cell, it can communicate with all the nodes available in that cell.

This assumption is reasonable if we take into account that:

1 low power wireless technologies like 802.15.4 and Bluetooth Low Energy have a coverage range in the order of 10m 30m in free space (although it depends greatly on several factors, such as antenna and operating frequency)

2 considering square coverage areas is actually a pejorative assumption, as we are clearly ignoring possible overlapping

3 the number of sensors considered is actually reasonable if we think of real possible applications (e.g. sensor for traffic lights, street lights, garbage cans, storm drains, etc.).

Around 6% of the grid cells can never be reached by the taxi cabs. According to visual inspections, these areas usually correspond to areas where the taxi cabs are not allowed to enter, such as public gardens, rivers, cemeteries, rail stations, and big private areas.

For vehicle-to-vehicle and vehicle-to-hotspot communications, we assume a coverage range of 250m that we consider a reasonable value for the 802.11a/b and 802.11p standards.

Trace acquisition and filtering

Each driver has a tablet device with the Android OS and an app that sends the GPS position every 7 seconds towards a server.
On the application side, the position is updated using the *getLastKnownLocation* method of the *LocationManager* Android object and is filtered against its precision, using the *getAccuracy()* function. This function returns the estimated accuracy in meters with 68% of probability. A sample is accepted only if its accuracy is less than 20m and discarded otherwise.

We subsequently filter the collected traces to mitigate some localization errors. In particular, analysing the traces we notice some "oddities" that we recognize because the distance between two subsequent points is greater than 125m which corresponds to a speed above 64km/h, a reasonable upper bound considering that the downtown speed limit is 50km/h. These oddities usually occur when the drivers are in a part of the city where the GPS service quality is poor (e.g. tunnels, tall buildings, etc.). In these cases we distinguish the duration of the anomaly: less than or greater than 42s (i.e. 6 points). In the former case we simply discard the "bad" samples. In the latter case (and if the anomaly does not last too long, i.e. less than 8 minutes), we correct the trace by introducing artificial samples according to the short path between the endpoints of the gap. For this task we use the Open Streetmap database. Finally, if the gap is greater than 8 minutes, we consider it a service interruption (lunch break, end of shift, etc.) and take no action.

After the data filtering described above, the position gathered by the android devices are assumed to be deterministically exact.

Statistical parameters of the evaluated traces

We took a bigger sample of five months (October 2013 to February 2014) to derive the statistical parameters resulting from the traces; the most important of which are related to speed and coverage.

Speed: During the day, taxi drivers can either move while serving customers (we call these movements "rides") or stay in parking lots if they are idle. To obtain the driving speed, we analyse 37327 rides (also outside the reference area) and obtain an average car speed of 31.9 km/h. The mean waiting time is 600s. The CDF of the average speed for each ride is shown in figure 2. As we can see, there are few cabs that move with an average speed greater than 50 km/h. This is a reasonable average speed considering stop times (e.g. traffic lights) and traffic congestion for the urban scenario presented. We point out that the speed limit in downtown is 50km/h. To quantify the time a node has to communicate with a cab, we analyse the average speed as seen by each cell. The empirical CDF of that measurement is shown in Figure 3 where we can see that more than 95% of the cells are crossed with an average speed less than 60km/h implying an average permanence of 2.4s in 95% of the cells, more than enough to allow simple data communication.

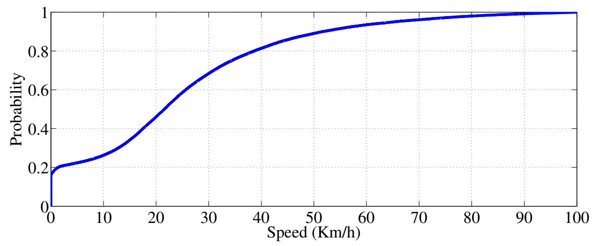


Figure 2: CDF of the average speed for each considered ride

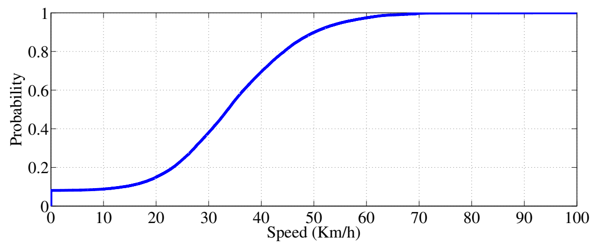
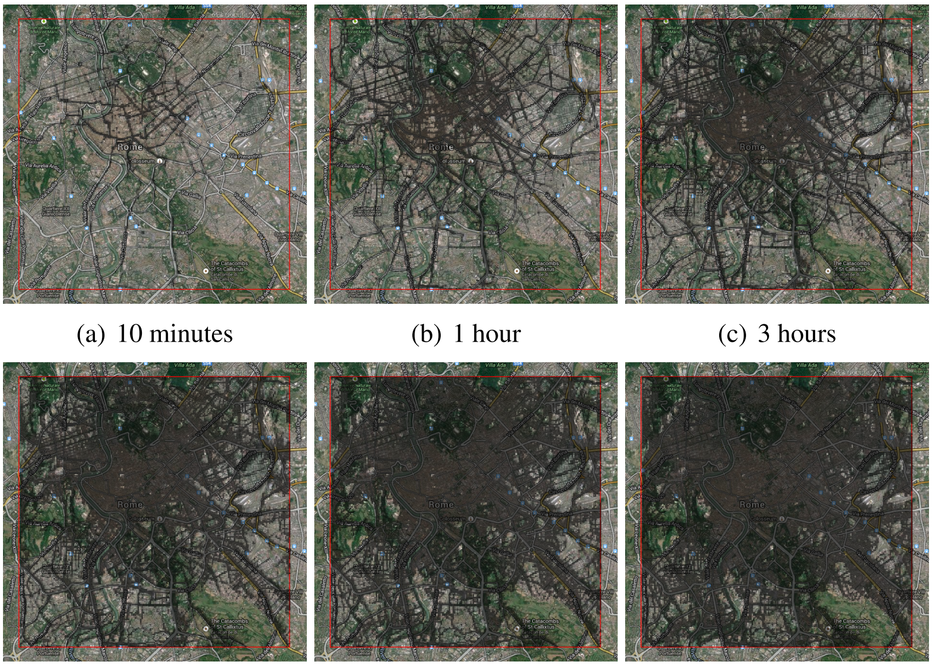


Figure 3: CDF of the average speed on each grid cell

Coverage: Figure 4 shows the presence of cars in the different parts of the reference area. In particular, for each cell, the overlay shows the average probability that there is at least one car in a reference period of 10 minutes up to 24 hours. If we consider the 6 hour period (figure 4d) we have a high probability that the cabs cover almost the entire area except for some zones not covered by the road network. When we consider a smaller reference period, 1 hour (figure 4b), the probability is greater than 0.8 only in the very center of the area and along the main roads. If we further shorten the period to 10 minutes, as represented in figure 4a, most of the area has a probability less than 0.2. To provide a baseline, figure 5 shows the same distribution for a synthetic trace where cars move according to a random waypoint mobility model, whose parameters (wait time and speed) are set to the same average values of the real traces. Comparing figures 4c and 5, we see how different the random waypoint model is from



(a) 10 minutes (b) 1 hour (c) 3 hours

(d) 6 hours (e) 12 hours (f) 24 hours

Figure 4: Probability that at least one car enters a cell varying the reference period using real traces. The grey scale overlay represents the probability from 0 to 1 with steps of 0.1: a darker area means a greater probability

reality: in fact, a random waypoint produces an almost complete coverage of the city after just half hour.

Validity

Throughout this paper we often present empirical probability distributions averaged for a month. It is worth asking if this data is a representative sample and if the analysis has a more general validity with respect to the analysed case. Indeed, statistics related to the taxi service vary dramatically according to a plethora of different factors such as the specific street topology of the city, the traffic, the weather, the month of the year, day or night time, working or non-working day, and many others. For instance, the demand of taxi cabs in Rome almost doubles during the summer with respect to the winter because of tourism. Notwithstanding, if we take a closer look at the statistical variability for a given city and for the considered sample, we notice some recurrent patterns.

Figure 6 compares the percentage of covered cells for a given time, considering a measurement period of one non-workday, one workday, and the weekly

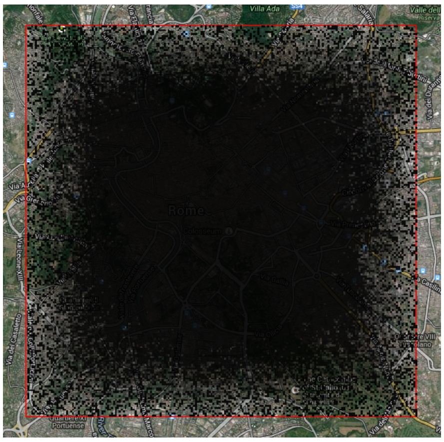


Figure 5: Probability that at least one car enters a cell in 30 minutes using a random waypoint mobility model

average. As we can see, for this performance metric, the differences are not very relevant, but if we take two working days (figure 7) the gap between the two curves becomes negligible. This calls for a *statistical guarantee* that we can provide to the proposed service model by considering an appropriate data analysis that must take into account only a few macroscopic factors, such as the average behaviour of a taxi fleet during summer/winter or work/non-work days.

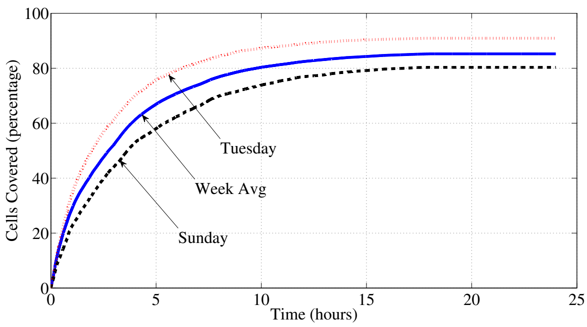


Figure 6: Percentage of covered cells for a given time, varying the measurement interval

Discussion

As a final remark, we point out that we limit our analysis to the downtown area: one could object that this scenario will not be applicable in rural or scarcely populated metropolitan areas. This is undoubtedly true. However, we consider

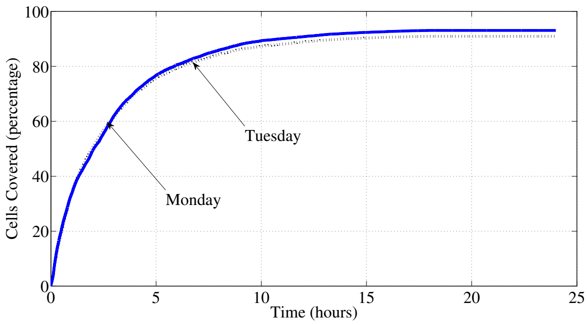


Figure 7: Percentage of covered cells for a given time, for two different working days

that the places (and duration) where more vehicles move, often correspond to the places where more data is produced/consumed and where communication is (relatively) more urgent. This consideration holds also if we look at the different zones of the same metropolitan area. Moreover, if we consider for instance special events like local fairs or public performances, we find that a greater number of taxi cabs in that zone corresponds to an increased need of communication among "things"; for instance to signal the trash can load levels so they are filled more rapidly. The result is that the oblivious data mule solution could also provide a kind of automatic adaptation to bring more capacity where and when it is needed.

4. System performance

We assess the performance of data dissemination and data gathering services using a period of one month from the real traces presented above. In the last part of this section we analyse some optimizations of energy requirements for such services and provide some insights on the trade-off between performance and energy requirements.

4.1. Data dissemination

To show the performance of the data dissemination process, we imagine that new data will arrive at a given time and need to be passed on to some nodes in the city. This data could be for instance an update of the interval of time specifying when street lamps should turn on and off, or a firmware update to reconfigure outdoor advertisements and road signs. To assess the performance of this system we imagine two different scenarios:

- **Online**: All vehicles are connected to the internet so all cars get the data as soon as it is available and the diffusion process to the nodes starts immediately. In this case users need to be cooperative in the sense that they should share their internet connection, but this will not influence their paths.

- **Hotspot**: There are some online hotspots so only when a car enters coverage range can it get the data we want to diffuse. Once the car gets the data, a simple epidemic routing algorithm will come into play and the data is passed to every other taxi cab that passes within 250m from a vehicle carrying the data. We simulate the presence of hotspots in four places in the city corresponding to four important taxi parking lots. This solution does not require internet connection sharing among drivers.

The simulation was realised with a software simulator specifically implemented for this work. We underline that both scenarios assume the following: (i) no particular propagation model for the wireless communication between sensors and vehicles was considered, (ii) to remain independent from the particular sensor network technology, the pairing delay was ignored.

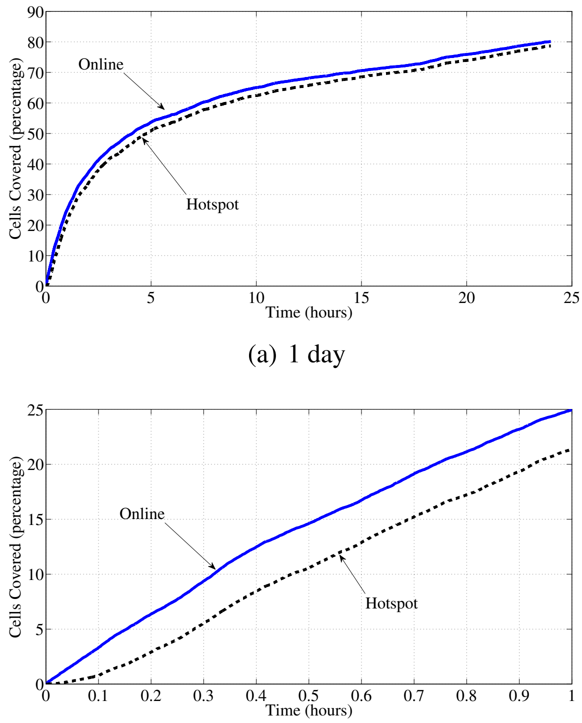
We remark that the delay achieved in the latter case by the epidemic diffusion represents a lower bound on the achievable delay. However, as the data grows, it could present severe scalability issues. Different DTN routing protocols could attain similar results without incurring in a similar data explosion, for instance by recurring to utility metrics and controlled replication (e.g. [9]) although an accurate analysis of these routing strategies is beyond the scope of this paper.

We simulate 4 different data diffusion occurences starting at different times of the day (6.00 a.m., 10.00 a.m., 2.00 p.m., and 6.00 p.m.) for every one of the 30 days. In figure 8 we plot the probability that a cell is reached by the data before a given time to provide insight on the *data diffusion delay*. As we can see, in 24 hours we can provide a coverage of 80% of the cells, whereas the difference in terms of delay among the two scenarios is not so relevant in the long run.

4.2. Data gathering

In the data gathering scenario we are interested in collecting a set of data generated by the smart city nodes. This scenario differs from the previous one because we have *multiple and disparate data*, generated by each cell whereas in the data diffusion scenario we have a single data message to be diffused to all cells.

In particular we are interested in the gathering performance in terms of *data retrieval delay* that is the time elapsed between when the data is produced inside



(a) 1 day

(b) 1 hour

Figure 8: Number of reached cells varying the diffusion time

the cell and when it is available online.
We consider the online scenarios presented above and measure the time between when the data is produced by a cell and when it is collected by a car. We repeated each experiment 30 times for all the 40,000 cells. Figure 9 shows the empirical cdf of the average retrieval delay. In the figure we also present only the values corresponding to the cells that can be reached by data mules in the reference month (94%). As we can see, more than half the data produced is gathered within one day, and after 2 days more than 90% are successful collections.

Even if from a strictly numerical point of view the data gathering analysis proves that almost complete city coverage with a limited number of vehicles is possible. It is important to note that we are not considering the temporal validity of the disseminated data, which might expire after a time window that depends on the particular application. Such an analysis is beyond the scope of this work.

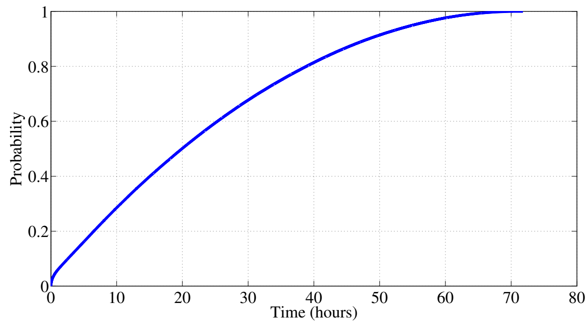


Figure 9: CDF of the data retrieval delay

4.3. Energy consumption optimization

To further improve energy saving in the deployed nodes, we present some optimizations with their performance results. All optimizations rely on the assumption that nodes could move from an active state, in which they are available for data transmission/reception, and a sleeping state in which they are not. The ratio of power consumption of these two states, usually in the order of 10^5 with current technology, justifies the need of such optimization. This assumption conforms to the most common energy models available in literature, even the ones that minimize wireless sensor life expectance [10].

Duty cycling

In the first scenario we introduce a constant **duty cycle** of the nodes so that they periodically cycle between two states ON and OFF for times that are respectively T_{ON} and T_{OFF}. To uniquely determine the behaviour of the duty cycle $d = T_{ON}/(T_{ON} + T_{OFF})$, we set $T_{ON} = 0.1s$, an interval of time that we consider sufficient to exchange small data messages between a node and the nearby data mule. Then we consider a successful data transmission if a car is in a cell during an ON period.

In figure 10 we plot the retrieval delay varying the duty cycle, considering only the online scenario for clarity. In the figure we present only the values corresponding to the cells that can be reached by data mules in the reference month (94%). For duty cycle values greater than 10% we have the same average delay of 9 hours. Indeed, $d \geq 0.1$ means the node is active every second and due to the granularity of the real trace, it is equal to having no duty cycle. If we decrease the duty cycle to 1% we double the delay that passes from 20 to 40 hours on average, obtaining a reduction of $1/10$ of the energy required by a single node.

This figure clearly demonstrates how we can balance communication delay with required energy.

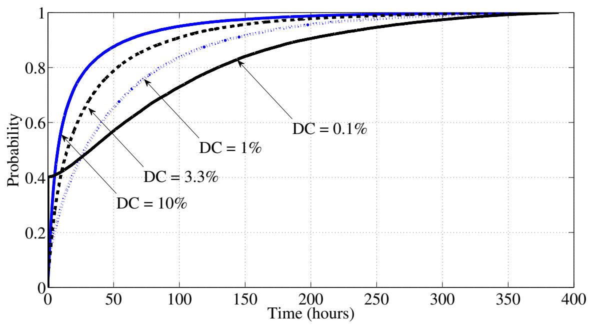


Figure 10: CDF of the retrieval delay, varying the duty cycle of the nodes

Windowing

In the second scenario, we consider a continuous interval of time (window) in which nodes are active after the data is produced. Figure 11 shows the cdf of the average retrieval delay varying the window. This scenario is interesting if we consider that data and energy are produced at the same time. For instance, by throwing garbage in a can we can produce a small amount of energy that is enough for a node to be active for a given period. The analysis shows that 73% of the nodes can be reached if they stay active for only 24 hours. An activity window of 24 hours seems a reasonable choice since after such a period the week average cell coverage is about 85% (as showed in Figure 6).

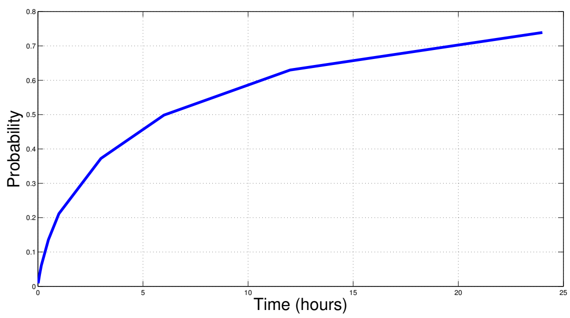


Figure 11: Average retrieval probability, varying the length of the active period (window) of the nodes. We highlight a slope change of around 6 hours for all possible duty circles which is reasonable and is related to the cell coverage time shown in Figure 4

Combining the two approaches

Figure 12 shows the combination of the two presented approaches. In this case, nodes are available for communication with data mules in their vicinity only if i) they are in their active period of the duty cycle, and ii) for up to *windows length* hours after data generation. As we can see, both factors dramatically impact the percentage of cells covered by the data gathering service. However, if we look at energy consumption, it is by introducing duty cycles that we can achieve consistent savings with relatively small performance degradation.

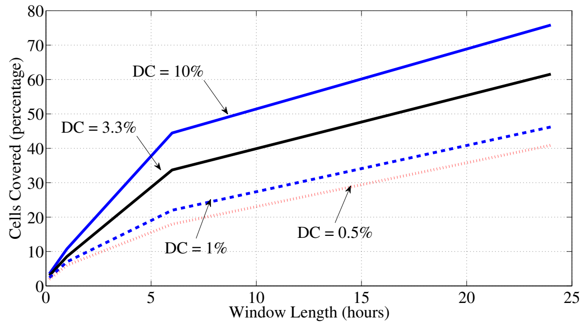


Figure 12: Average retrieval probability, varying the length of the active period (window) and the duty cycles of the nodes

5. Related works

Mobile nodes

The idea of mobile nodes configured to store, carry, and forward information, was first considered in the research field of sparse/partitioned ad-hoc networking, where a node's mobility is intrinsic in the network [11].

Subsequently, mobile nodes have also been considered for sparse wireless sensor networks to cope with the limitation in the energy budget, thus allowing the use of short range communication technologies. In this regard, all literature can be classified according to the assumptions made on node mobility. In [2] the mobile nodes (namely *data mules*) are assumed to move randomly in a two-dimensional square area to evaluate the basic performance of a data gathering application. The same conclusions regarding the performance of the sensor/data mule connections (for both discovery and data transfer), in terms of energy efficiency, are reported in [12] or in [13]. Other works assume that node mobility is **predictable** or even **controllable**. In [14] *message ferries* are used for proactive data delivery in sparse networks exploiting the predictability of node movement. Some works such as [4], [5], and [6] focused on the optimization of the complex problem of controlling/planning node mobility with optimal or suboptimal techniques.

As briefly mentioned above, we consider uncontrolled vehicles moving around the city as mobile nodes. We demonstrate, in the remainder of this paper, the typical movement of a car is a key aspect, since it is not uniform - neither spatially nor in velocity.

VANET and Smart Cities

The integration of VANETs in smart cities is considered under several different aspects. The VANET can be seen as an opportunity for sensing data from the surrounding environment in order to assist navigation, pollution control, and traffic management as grasped by [15]. Similarly, some works considered the interaction that the VANET can have with the surrounding environment through wireless sensor networks. The work proposed in [16] derives the optimal senor placement along the road to obtain full coverage for navigation support. The work in [17] describes and evaluates a driving safety application exploiting an integrated roadside sensor network.

In [18] the interaction between domestic WiFi networks and VANET technologies is considered. In this work cars are user to exchange information with a given access point using the beacon channel.

Other works consider the urban environment only as a constraint to the mobility model as in various optimization or architecture works (for example [19] and [20]).

Synthetic traces and mobility models for VANET

The assessment of VANET protocols is typically performed combining two different evaluation approaches: i) generating vehicle movements in the interested area using a **mobility simulator**; ii) simulating communication among nearby vehicles with a **network simulator** (e.g. with NS2). In the most simple case, VANET protocols are validated using only synthetic movement traces (produced by mobility simulators). In this case, cars can exchange data if their distance is below a given threshold, which depends on many factors and corresponds to the specific wireless technology, such as 802.11b/g/a or the more recent 802.11p [21]. A realistic mobility model should include several aspects of real mobility (such as one way streets, traffic lights, obstacles, weather conditions, drivers' behavior, etc.) that are complex to model and to take into account. For this reason, several simulators have been developed.

IMPORTANT framework [22] is one of the earlier simulators for MANET protocols. It combines several mobility models including the naive random waypoint (RWP) algorithm, RPGM model for group mobility, the Freeway mobility model that considers a single high speed street, and the Manhattan mobility model where vehicles move on a grid. From any of these models, the simulator can produce a connectivity graph that can be used to assess the performances of different routing protocols. As pointed out by the authors, RWP presents several limitations because it neglects the temporal, the spatial, and the geographical dependencies of the vehicles. Notwithstanding, it is still the "most commonly used mobility model in the MANET research community".

VanetMobiSim [23] [24] is an extension to the CANU Mobility Simulation Environment (CanuMobiSim), for vehicular mobility. VanetMobiSim can import maps from the US Census Bureau TIGER/Line database, or randomly generate them using Voronoi tesselation. This simulator goes in the direction of simulating a near-to-reality scenario, hence it offers support for many mobility features such as multi-lane roads, separate directional flows, differentiated speed constraints and traffic signs at intersections. For this reason, it must be tuned using several complex parameters, often hard to quantify.

TRANSIM [25] is an integrated set of tools developed to analyze regional transportation systems. It integrates several aspects of mobility, including the simulated behavior of public transportation and many algorithms to reproduce the regional population to match real population demography. It uses cellular automata to simulate interactions among vehicles. Another alternative is SUMO [26], an open source mobility simulator that uses a Gipps-model to simulate the main features of traffic flow, while taking into account a wide range of editable features such as traffic lights etc.

Due to the complexity of these simulators, researchers often resort to simpler models such as the naive Random Waypoint, or Random Waypoint City Model [27] that includes city maps, or STRAW [28], that adds vehicular congestion and simplified traffic control mechanisms. All these models could be enriched by adding more parameters (see [29] for a complete survey), paying the price of adding greater intricacy to the already complex environment.

In [30] the authors analyze the different impacts of the features introduced to simulate urban mobility and draw the conclusion that some features, such as waiting at intersections, affects simulation results more significantly than others, such as multiple lanes or coordinated traffic lights.

Real traces for VANET

Some works use real traces [31] to study the storage capability of VANET [32] or the dynamics of network topologies [33]. Usually, these real traces are provided by tracking public transportation vehicles. Among them, the traces obtained by tracking taxi cabs are particularly important with respect to the ones obtained by tracking buses, because the former better explore the status of city streets. In this field, the most used public domain traces available in literature are the GPS traces of 533 taxis collected in 20 days in the San Francisco Bay area, USA [34] and the traces of 13,799 taxi cabs collected in 9 days in the city of Shenzhen, China.

In this paper we present an analysis in the city of Rome that presents a different topology compared to Shenzhen and the San Francisco Bay area (near) grid topology. Another alternative is to use other traces available from public services (eg. NOKIA Sports Tracker) and acquired by logging mobile phone positions. The down side is that these traces present only some time segments and are therefore not suitable for an extensive analysis. Moreover, there is no guarantee that the samples in the trace refer to vehicles and not pedestrians, trains, etc.

In [35], authors infer traffic volumes though data collected by a taxi fleet, explorating current patterns and conditions both from historical and real-time collected data. To this aim, authors performed a large scale experiment with more than 16000 taxi fleet in the Singapore city. Possible usage are for traffic visualization, analysis and urban planning. Differently from this work, authors of [35] focused more on traffic visualization, analysis and urban planning rather than opportunistic communication with the sensors of a smart city, as we did. However, that work confirm that taxi distributions are rapidly converging, that is, taxi will randomly visit locations based on the random nature of the passengers' destinations. This is a further confirmation of the proposed taxi-based application scenario represents a good choice.

CarTel is a MIT project that helps applications easily collect, process, deliver, analyze, and visualize data from sensors located on mobile units (mobile phones and in-car embedded devices). In particular, this project has a network subsystem with a component called CafNet (carry and forward network) [36] that realize a delay-tolerant networking for mobile data muling. More in details, it uses a combination of WiFi, Bluetooth, and cellular connectivity, using whatever mode is available and working well at any time. Our project instead has the different goal to validate the feasibility of a sensor network data collection scenario.

More recently, [37] uses a 450 GB dataset of 14000 taxicabs to infer passenger arriving moments by interactions of vacant taxicabs, and infer passenger demand by a customized online training with both historical and real-time data. Once again the focus is different as a more focused on a big data methodology and modeling.

Due to the road topology of Rome, and its traffic conditions, which are far from either the realistic mobility traces in modern simulators and the real mobility traces acquired in cities like San Francisco and Shenzen (which are compatible with simple grid topology assumptions), the evaluated scenario and the proposed traces extend the testing cases available for validating the performance of the VANET protocols.

6. Conclusions and future works

In this work we assessed the feasibility of a data mule application in which about 320 taxi cabs are used as mobile nodes to gather and disseminate data from/to a large number of sensors deployed in the city of Rome. We showed that we can provide some statistical guarantee to enhance many specific services, given that communication is provided almost "for free", boosting the IoT with new capacity. To the best of our knowledge, this work is the first to ever validate data mule applications with real world mobility traces.

More specifically, we identified a set of performance metrics and evaluated different service scenarios for data dissemination and gathering from data mules to IoT nodes and vice-versa. We based our analyses on a real experiment that was conducted in Rome, Italy, representing the longest (6 months) data collection available with the finest sample rate (7 sec). We further focus on energy saving strategies by analyzing energy optimization techniques to show how performance metrics are affected by windowing and adaptive duty cycles.

Possible future works include: (i) extending the experiment campaign to other cities to understand how different city topologies affect system performance; (ii) improving the statistical analysis by showing the variability of the statistical guarantees provided by this work; (iii) considering mobile node fleets with different sizes, comparing the results and analyzing how this can affect overall performance; (iv) implementing an actual application for commercial tablets equipped with low energy bluetooth and running a real world trial to assess the system performance on the field; (v) generating experimental evidence stressing data communication between sensors and mules in different weather conditions and considering the pairing delay.

References

[1] T. Tanaka, T. Suzuki, K. Kurihara, Energy harvesting technology for maintenance-free sensors, FUJITSU Sci. Tech. J 50 (1) (2014) 93–100.

[2] R. Shah, S. Roy, S. Jain, W. Brunette, Data mules: modeling a three-tier architecture for sparse sensor networks, in: Sensor Network Protocols and Applications, 2003. Proceedings of the First IEEE. 2003 IEEE International Workshop on, 2003, pp. 30–41. doi:10.1109/SNPA.2003.1203354.

[3] R. A. Gorcitz, Y. Jarma, P. Spathis, M. Dias De Amorim, R. Wakikawa, J. Whitbeck, V. Conan, S. Fdida, Vehicular Carriers for Big Data Transfers (Poster), in: 2012 IEEE Vehicular Networking Conference (VNC), Seoul, Corée, République De, 2012, pp. 1–8. URL `http://hal.archives-ouvertes.fr/hal-00739361`

[4] A. Kansal, M. Rahimi, D. Estrin, W. Kaiser, G. Pottie, M. Srivastava, Controlled mobility for sustainable wireless sensor networks, in: Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on, 2004, pp. 1–6. doi:10.1109/SAHCN.2004.1381896.

[5] A. Somasundara, A. Kansal, D. Jea, D. Estrin, M. Srivastava, Controllably mobile infrastructure for low energy embedded networks, Mobile Computing, IEEE Transactions on 5 (8) (2006) 958–973. doi:10.1109/TMC.2006.109.

[6] R. Sugihara, R. K. Gupta, Path planning of data mules in sensor networks, ACM Trans. Sen. Netw. 8 (1) (2011) 1:1–1:27. doi:10.1145/1993042.1993043. URL `http://doi.acm.org/10.1145/1993042.1993043`

[7] F. L. Piccolo, D. Battaglino, L. Bracciale, M. Di Filippo, A. Bragagnini, M. S. Turolla, N. B. Melazzi, Towards fully ip-enabled ieee 802.15. 4 lr-wpans, in: Sensor, Mesh and Ad Hoc Communications and Networks Workshops, 2009. SECON Workshops' 09. 6th Annual IEEE Communications Society Conference on, IEEE, 2009, pp. 1–3.

[8] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, A. Rabuffi, CRAWDAD dataset roma/taxi (v. 2014-07-17), Downloaded from http://crawdad.org/roma/taxi/20140717 (Jul. 2014). doi:10.15783/C7QC7M.

[9] T. Spyropoulos, K. Psounis, C. S. Raghavendra, Spray and wait: An efficient routing scheme for intermittently connected mobile networks, in: Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking, WDTN '05, ACM, New York, NY, USA, 2005, pp. 252–259. doi:10.1145/1080139.1080143. URL `http://doi.acm.org/10.1145/1080139.1080143`

[10] M. N. Halgamuge, M. Zukerman, K. Ramamohanarao, H. L. Vu, An estimation of sensor energy consumption, Progress In Electromagnetics Research B.

[11] Z. Zhang, Routing in intermittently connected mobile ad hoc networks and delay tolerant networks: overview and challenges, Communications Surveys & Tutorials, IEEE 8 (1) (2006) 24–37.

[12] G. Anastasi, M. Conti, M. Di Francesco, Data collection in sensor networks with data mules: An integrated simulation analysis, in: Computers and Communications, 2008. ISCC 2008. IEEE Symposium on, 2008, pp. 1096–1102. doi:10.1109/ISCC.2008.4625629.

[13] G. Anastasi, M. Conti, E. Monaldi, A. Passarella, An adaptive data-transfer protocol for sensor networks with data mules, in: World of Wireless, Mobile and Multimedia Networks, 2007. WoWMoM 2007. IEEE International Symposium on a, 2007, pp. 1–8. doi:10.1109/WOWMOM.2007.4351776.

[14] W. Zhao, M. Ammar, E. Zegura, A message ferrying approach for data delivery in sparse

mobile ad hoc networks, in: Proceedings of the 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc '04, ACM, New York, NY, USA, 2004, pp. 187–198. doi:10.1145/989459.989483.
URL http://doi.acm.org/10.1145/989459.989483

[15] M. Gerla, E.-K. Lee, G. Pau, U. Lee, Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds, in: Internet of Things (WF-IoT), 2014 IEEE World Forum on, IEEE, 2014, pp. 241–246.

[16] M. Rebai, L. Khoukhi, H. Snoussi, F. Hnaien, Optimal placement in hybrid vanets-sensors networks, in: Wireless Advanced (WiAd), 2012, 2012, pp. 54–57. doi:10.1109/WiAd.2012.6296567.

[17] H. Qin, Z. Li, Y. Wang, X. Lu, W. Zhang, G. Wang, An integrated network of roadside sensors and vehicles for driving safety: Concept, design and experiments, in: Pervasive Computing and Communications (PerCom), 2010 IEEE International Conference on, 2010, pp. 79–87. doi:10.1109/PERCOM.2010.5466988.

[18] S. Schnaufer, S. Kopf, H. Lemelson, W. Effelsberg, Beacon-based short message exchange in an inner city environment, in: Ad Hoc Networking Workshop (Med-Hoc-Net), 2010 The 9th IFIP Annual Mediterranean, 2010, pp. 1–8. doi:10.1109/MEDHOCNET.2010.5546868.

[19] J. Luo, X. Gu, T. Zhao, W. Yan, Mi-vanet: A new mobile infrastructure based vanet architecture for urban environment, in: Vehicular Technology Conference Fall (VTC 2010-Fall), 2010 IEEE 72nd, 2010, pp. 1–5. doi:10.1109/VETECF.2010.5594510.

[20] R. Akamatsu, M. Suzuki, T. Okamoto, K. Hara, H. Shigeno, Adaptive delay-based geocast protocol for data dissemination in urban vanet, in: Mobile Computing and Ubiquitous Networking (ICMU), 2014 Seventh International Conference on, 2014, pp. 141–146. doi:10.1109/ICMU.2014.6799085.

[21] C. Campolo, A. Molinaro, On vehicle-to-roadside communications in 802.11p/wave vanets, in: Wireless Communications and Networking Conference (WCNC), 2011 IEEE, 2011, pp. 1010–1015. doi:10.1109/WCNC.2011.5779273.

[22] F. Bai, N. Sadagopan, A. Helmy, The important framework for analyzing the impact of mobility on performance of routing protocols for adhoc networks, AdHoc Networks Journal 1 (2003) 383–403.

[23] J. Härri, M. Fiore, F. Filali, C. Bonnet, Vanetmobisim.
URL http://vanet.eurecom.fr/

[24] J. Härri, F. Filali, C. Bonnet, M. Fiore, Vanetmobisim: Generating realistic mobility patterns for vanets, in: Proceedings of the 3rd International Workshop on Vehicular Ad Hoc Networks, VANET '06, ACM, New York, NY, USA, 2006, pp. 96–97. doi:10.1145/1161064.1161084.
URL http://doi.acm.org/10.1145/1161064.1161084

[25] L. Smith, R. Beckman, D. Anson, K. Nagel, M. Williams, TRANSIMS: Transportation analysis and simulation system, in: Conference: 5. National transportation planning methods applications conference, Seattle, WA (United States), 17-21 Apr 1995, 1995.

[26] D. Krajzewicz, J. Erdmann, M. Behrisch, L. Bieker, Recent development and applications of SUMO - Simulation of Urban MObility, International Journal On Advances in Systems and Measurements 5 (3&4) (2012) 128–138.
URL http://elib.dlr.de/80483/

[27] J. Kraaier, U. Killat, -the random waypoint city model-: user distribution in a street-based

mobility model for wireless network simulations, in: Proceedings of the 3rd ACM international workshop on Wireless mobile applications and services on WLAN hotspots, ACM, 2005, pp. 100–103.

[28] D. R. Choffnes, F. E. Bustamante, An integrated mobility and traffic model for vehicular wireless networks, in: Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks, ACM, 2005, pp. 69–78.

[29] E. Spaho, L. Barolli, G. Mino, F. Xhafa, V. Kolici, Vanet simulators: A survey on mobility and routing protocols, in: Broadband and Wireless Computing, Communication and Applications (BWCCA), 2011 International Conference on, 2011, pp. 1–10. doi:10.1109/BWCCA.2011.11.

[30] A. Mahajan, N. Potnis, K. Gopalan, A.-I. A. Wang, Urban mobility models for vanets, in: IN PROC. OF 2ND WORKSHOP ON NEXT GENERATION WIRELESS NETWORKS, 2006.

[31] M. Piorkowski, N. Sarafijanovic-Djukic, M. Grossglauser, A parsimonious model of mobile partitioned networks with clustering, in: Communication Systems and Networks and Workshops, 2009. COMSNETS 2009. First International, 2009, pp. 1–10. doi:10.1109/COMSNETS.2009.4808865.

[32] B. Liu, B. Khorashadi, D. Ghosal, C.-N. Chuah, H. M. Zhang, Analysis of the information storage capability of vanet for highway and city traffic, Transportation Research Part C: Emerging Technologies 23 (2012) 68–84.

[33] Y. Chen, M. Xu, Y. Gu, P. Li, X. Cheng, Understanding topology evolving of {VANETs} from taxi traces, Advanced Science and Technology Letters 42 (Mobile and Wireless) (2013) 13–17.

[34] [link].
URL `http://crawdad.cs.dartmouth.edu/`

[35] J. Aslam, S. Lim, X. Pan, D. Rus, City-scale traffic estimation from a roving sensor network, in: Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, SenSys '12, ACM, New York, NY, USA, 2012, pp. 141–154. doi:10.1145/2426656.2426671.
URL `http://doi.acm.org/10.1145/2426656.2426671`

[36] K. W. Chen, Cafnet: A carry-and-forward delay-tolerant network, Ph.D. thesis, Massachusetts Institute of Technology (2007).

[37] D. Zhang, T. He, S. Lin, S. Munir, J. Stankovic, Dmodel: Online taxicab demand model from big sensor data in a roving sensor network, in: Big Data (BigData Congress), 2014 IEEE International Congress on, 2014, pp. 152–159. doi:10.1109/BigData.Congress.2014.30.