

A Bus-based Opportunistic Sensing Network

Pedro Henrique Cruz Caminha

Advisors: Luís Henrique Maciel Kosmowski Costa and Rodrigo de Souza Couto

¹Programa de Engenharia Elétrica – Universidade Federal do Rio de Janeiro (UFRJ)

{cruz, luish, rodrigo}@gta.ufrj.br

***Abstract.** Smart city applications need data about the city, and this data must follow specific requirements. Two of these requirements are the maximum delivery delay and the minimum measurement frequency. Using buses to gather data and bus stops as gateways can be cost-effective, but data might not fit the application requirements. In this thesis, we present a model to minimize the delay of data delivery, a metric to estimate the coverage, and a prototype of the nodes of such a network. We use GPS data from the bus fleet of Rio de Janeiro to show that it is possible to cover a significant part of the city, fulfilling application requirements specified by the smart city literature.*

1. Problem Statement and Contributions

Granting mobility to sensors mobility is a profitable strategy to gather data and build smart city applications. Under the Internet of Things (IoT) paradigm, we consider in this work that urban buses carry inexpensive and resource-limited sensors. These IoT devices collect data and send it to more powerful devices located in the fog, following the current three-tier hierarchical cloud computing model [Li et al. 2016]. The fog pre-processes data at the edge of the network, sending it to the cloud through the Internet. Finally, the cloud processes and makes data available to applications. The city is then covered by a Mobile Wireless Sensor Network (MWSN), which presents two main trade-offs. First, sensors move through the city, enlarging their individual coverage and consequently reducing the amount of IoT devices to cover the same area as a static approach [Liu et al. 2005]. The cost to that is that each region of the city is not covered the whole time. The second trade-off comes from the fact that mobile sensors can take advantage of opportunistic communication. In this situation, there is no need for a network covering the whole city: sensors collect data and send it when they approach gateways [Ekici et al. 2006]. Since buses collect data and store it until a connection is possible, there is a delay added to the data. Given this reasoning, we make three contributions:

- a method to minimize the delivery delay by the optimal positioning of gateways;
- a metric to define the coverage of the applications served by a bus-based MWSN;
- a prototype of SensingBus, a bus-based MWSN.

2. Related Work

Even though the literature presents works to minimize delays on static [Wong et al. 2004, Umer et al. 2016] and mobile [Ghafoor et al. 2014] Wireless Sensor Networks (WSNs), their focus is to optimize trajectories or packet routing. Optimizing trajectories is not possible on a mobile network composed of buses following fixed trajectories. Also, packet routing is not significant to the case we consider, in which buses deliver data directly to

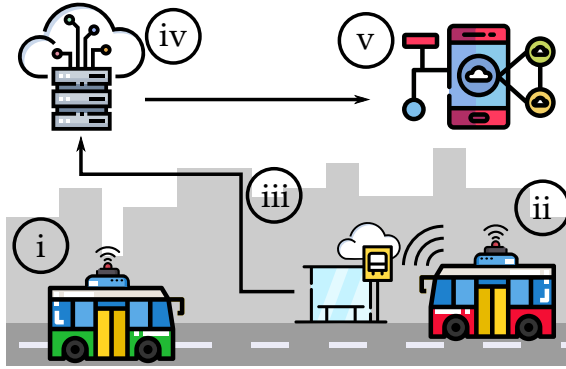


Figure 1. SensingBus overview.

Application	Tolerated delay (s)	Measurement frequency (day^{-1})
Waste management	1,800	24
Air quality monitoring	300	48
Noise monitoring	300	144

Table 1. Applications data requirements, according to [Zanella et al. 2014]

the fog nodes. It is also possible to find extensive work on the coverage of vehicle-based MWSNs [Ali and Dyo 2017, Gao et al. 2016, Zhao et al. 2013]. These metrics are suited for MWSNs that are constantly connected. In contrast, we propose a coverage metric that uses the streets covered by buses, observing the measurement frequency and the delivery delay as requirements of the targeted applications. Regarding the SensingBus, there are other bus-based MWSN prototypes in the literature [Zoysa et al. 2007, Gao et al. 2016, Marjovi et al. 2015, Sanchez et al. 2014]. These prototypes do not consider a fog level on their architecture. In this work, we implement the three-level fog architecture and perform experiments related to processing, communication, and mobility.

3. Main Assumptions

Our bus-based MWSN is depicted in Figure 1. In this network, buses collect and store data throughout their trajectory (i), delivering data to gateways placed into bus stops (ii). The gateways send the information to the cloud node (iii), that processes the data (iv) and provides it to the corresponding applications (v). In the present work, the gateways are fog nodes that receive data and also perform pre-processing, as we explore in Section 6.

To work efficiently, smart city applications have specific requirements on the data they receive. We select the requirements we believe our scenario impacts the most, the *maximum delivery delay* and the *minimum measurement frequency*. They mean, respectively, that applications must receive data about a region within a certain delay, and that the measurements about each region must be frequent. Otherwise, the application is not able to cover the region. Table 1 shows the tolerated delay and the measurement frequency of the applications we consider in this work, according to Zanella *et al.* [Zanella et al. 2014].

4. A Delay Optimization Model for Bus-based MWSNs

In the MWSN we consider, buses gather data all the time, but can only deliver it when a connection to a fog node is possible. We assume that every time a bus gathers a piece of data, it delivers this piece to the next fog node in its bus path. We also assume that the fog nodes are located in the bus stops. This means that the time a bus takes between two bus stops is also the delay experienced by the data gathered between these bus stops. Throughout their paths, each bus makes contact with several bus stops, delivering all data

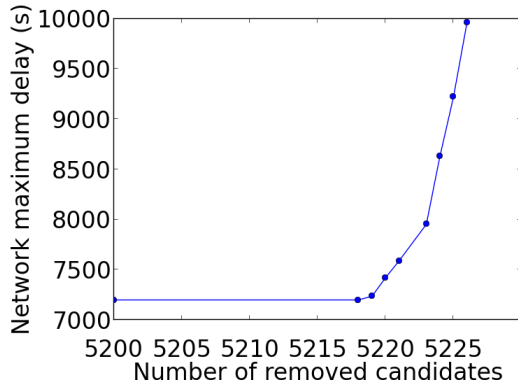


Figure 2. Network delay for the buses filtered at 7,200s for different budgets.

Budget (fog nodes)	Delay (s)	Execution time (s)
200	1786	1601
300	1786	1643
400	1786	1354
500	1786	1376
600	1786	1378
700	1786	1381

Table 2. Results of the fog node placement with reduced cardinality.

stored since the last contact. A bus stop can receive data from multiple buses, even from buses not serving passengers on that bus stop.

We assume that the cost might not allow the installation of fog nodes into all bus stops. In this situation, some fog nodes are removed, meaning that buses that pass by these stops need to deliver data to the next fog nodes in their paths. Since each bus stop receives data from buses that cover different paths, removing a bus stop can affect each bus differently. It is possible to choose the best bus stops to install fog nodes, minimizing the biggest delay suffered by any data in the network.

We model the optimal fog node placement as a p-center problem, a well-known facility location problem [Kariv and Hakimi 1979]. In our model, each bus leaving a bus stop is a demand and the facility candidates capable of serving the demand are the next bus stops in the bus path. The distance between a demand and a facility candidate is the time the bus takes to go from the first data acquisition until its delivery. A possible representation of this problem is as a bipartite graph, in which facility candidates and demands are the two sets of vertices, and the distances are the edges.

Obtaining a solution to the optimal fog node placement problem is computationally intensive, and instances to the size of a real city can be impractical. We develop an algorithm capable of finding suboptimal solutions in time $O(kmn)$, where k is the number of buses, m is the number of bus stops in the longest path of buses, and n is the number of bus stops of the city. The algorithm calculates a cost to the delay for removing each bus stop and removes the most expensive one. This procedure is repeated until the number of remaining bus stops obeys the budget. We use 30 artificial datasets with a known optimal solution and show that our algorithm, in every case, finds a solution that is no more than 10% distant from the optimal solution.

We collect a dataset with the GPS positions of 6,683 buses of Rio de Janeiro for 24 h and the positions of its 6,272 bus stops. We filter the dataset for inconsistencies, such as coordinates outside of Rio de Janeiro. We also eliminate from the dataset buses that have delays considered too big to serve the network. In the sequence, we run the algorithm for different budgets. Figure 2 illustrates the network delay we obtain for different budgets. In this case, we eliminate from the dataset all the buses that have at least one

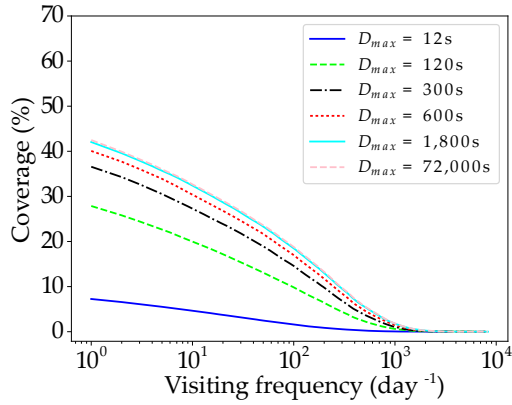


Figure 3. Coverage of Rio de Janeiro as a function of the visiting frequency, for different maximum delays.

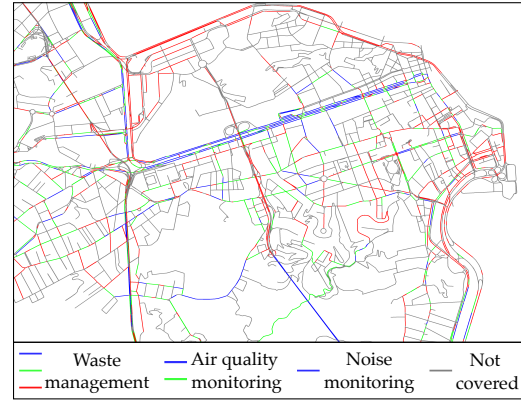


Figure 4. Coverage of the central region of Rio de Janeiro for different smart city applications.

delay bigger than 7,200 s in their paths (5,429 buses remaining). We present results for other filter values in the thesis [Cruz 2020].

We can use the solutions of our suboptimal algorithm to reduce the cardinality of the p-center problem. The reduction allows us to find optimal solutions to bigger instances of the problem. We do this by eliminating from the formulation all the delays that are bigger than the suboptimal maximum delay. In practice, this means eliminating from the graph all the edges that are heavier than the suboptimal value. We collect 24 h of GPS data from 116 buses and the coordinates of 744 bus stops of the South Zone of Rio de Janeiro. This dataset, when transformed into a p-center problem, generates a set of 1,122,629,838 edges. After we run the suboptimal algorithm for a budget of 120 bus stops, the number of edges is 393,157. The runs of the suboptimal algorithm take less than 3 s each. Since we directly use real data, our method considers all the traffic conditions within the dataset. Table 2 shows the delays and the execution times for different budgets.

5. Coverage of Bus-based MWSNs

The coverage is an important metric for a sensor network, indicating which regions a certain application can serve. We propose a simple coverage metric and then enrich this metric to consider the delays and the measurement frequency of each region.

Buses travel through the city and collect data about the streets they visit. A common representation of the street map of a city is a graph, in which vertices are points of interest (an intersection, for instance) and an edge between two vertices is a street that directly connects these two points. We call each edge a street segment. We then associate to each edge a weight, representing the size of the street segment. We define that a street segment e is covered if and only if any bus passes through e . We then define the coverage of a street as the sum of the weights of all covered street segments.

It is possible that budget restrictions exist, and thus not every bus will receive a piece of sensing equipment. Given this assumption, we formulate a problem that maximizes the coverage for a limited number of buses. We show that this problem is equivalent to the Maximal Coverage Location Problem [Church and Velle 1974]. We collect

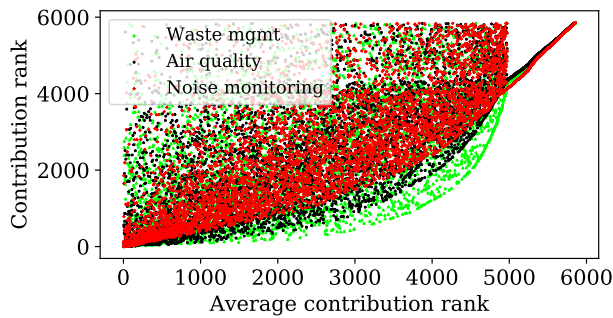


Figure 5. Bus contribution rank as a function of average contribution rank, for different applications.

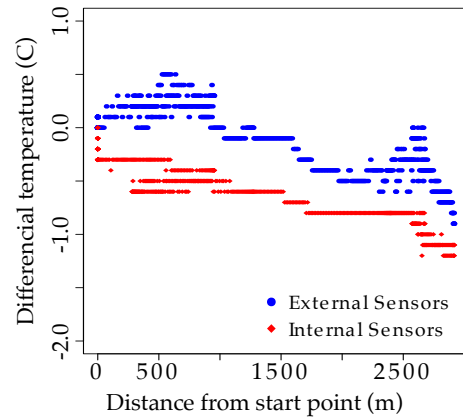


Figure 6. Temperature difference between measurements obtained inside and outside an acrylic case.

a dataset containing 5,496,878 GPS coordinates from 6,075 buses of Rio de Janeiro and enrich this dataset to reconstruct the routes of the buses in terms of street sections. We then instantiate the optimization problem to these buses, for different budgets. We show that 1,024 buses, or approximately 18% of the bus fleet, can cover at least 94% of the streets served by buses (which is about 50% of the total streets of the city).

5.1. A delay-aware coverage metric

The metric we propose in Section 5 is well suited for applications very tolerant to delays and low measurement frequencies. Nevertheless, some smart city applications have hard constraints on these requirements, as we show in Table 1. Hence, we also develop another metric, that takes into account the constraints of maximum delivery delay and minimum measurement frequency.

We consider that a street is visited once if a bus with sensing equipment passes through this street section and makes contact with a fog node before the maximum delivery delay of the targeted application. Then we say that this street section is covered if and only if it is visited with at least the minimum measurement frequency of the targeted application. The total coverage is the sum of the lengths of covered street sections, normalized by the sum of the lengths of all the street sections. Note that this coverage metric is application-specific.

We collect 29,155,221 GPS positions of 5,706 buses and enrich this dataset to reconstruct the routes of the buses in terms of street segments, but also to detect contacts between buses and bus stops. We then build an abacus, relating the coverage to the minimum measurement frequency, for different values of maximum delivery delay. Figure 3 illustrates this abacus. To understand the geographic coverage distribution, Figure 4 illustrates the coverage map of the targeted applications in the city center of Rio de Janeiro.

5.2. Bus coverage contribution

We also propose a metric for the coverage contribution of each bus. This metric consists of the parcel of coverage lost when the bus is not participating in the sensing tasks. We use the real data from Rio de Janeiro to rank the buses, for each application. Figure 5

shows the contribution of each bus for the applications in Table 1, as a function of the average contribution, obtained for the three applications. We also compare the rankings two by two, showing the low correlation between coverage contribution for two different applications, with a Kendall coefficient of 0.21 or less in every case. This shows that the contribution of each bus is highly correlated to the application the MSWN serves.

Understanding the delay-aware coverage metric and the bus coverage contribution is important to develop better bus-based MWSN's. Stakeholders can use the delay-aware coverage metric to understand the utility of the envisioned MWSN, while the bus coverage contribution can help deciding the participating buses and maintenance schedules.

6. Prototype of SensingBus Nodes

We develop and evaluate a prototype for the different nodes types of SensingBus, as a proof-of-concept. SensingBus uses the three-tier architecture mentioned in Section 1.

To develop a sensing node, we use an Arduino UNO, together with an ESP8266, a GPS and memory shield, and a sensor bank with temperature, light intensity, rain intensity, and barometric pressure sensors. We write software on Arduino to read the sensors and store data in the memory. We rewrite the ESP8266 firmware to connect to fog nodes and send the stored data. The equipment is protected by a case made of transparent acrylic. We perform tests to access the adequacy of the measurements under the presence of mobility and also to access the effect of the case. Figure 6 illustrates the difference in the temperature measured by a sensor inside the acrylic case and a sensor outside the acrylic case. This shows that casing can affect very precise measurements.

We also develop a fog node, capable of receiving data from the sensing node, pre-processing data, and sending it to the cloud. The pre-processing consists of checking for common inconsistencies and compressing data. The fog node is a Raspberry Pi model B with a WiFi dongle. Figure 7 shows a test setup with 20 sensing nodes and one fog node. In Figure 8, we can observe that the fog node can serve up to 20 sensing nodes without exhausting its memory and CPU resources. It is worth noting that the average throughput growth is not linear. This means that the throughput obtained by each sensing node drops as the number of sensing nodes increases, as a consequence of competition for the transmission medium.

7. Impact

We list below the published results of this thesis. They were granted two conference awards. If applicable, the publications are accompanied by acronyms: "QCC" and "QEIV" stand for the publication Qualis 2013-16 under "Ciência da Computação" or "Engenharias IV", whereas "IF" denotes its JCR'2019 impact factor:

1. **Cruz, P.**, Pinto Neto, J. B., Campista, M. E. M., and Costa, L. H. M. K. "On the Accuracy of Data Sensing In the Presence of Mobility." In **Int. Conference on Network of the Future**, Nov. 2016.
2. **Cruz, P.**, Couto, R. S., and Costa, L. H. M. K. "Um Algoritmo de Posicionamento de Pontos de Coleta para uma Rede de Sensores Baseada em Ônibus Urbanos." In **Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, May 2017. QCC: B2. – **Awarded honorable mention.**
3. **Cruz, P.**, Silva, F. F., Pacheco, R. G., Couto, R. S., Velloso, P. B., Campista, M. E. M., and Costa, L. H. M. K. "SensingBus: um Sistema de Sensoriamento Baseado em Ônibus Urbanos." In **Salão de Ferramentas do Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, May 2017. QCC: B2. – **Awarded best demo paper.**

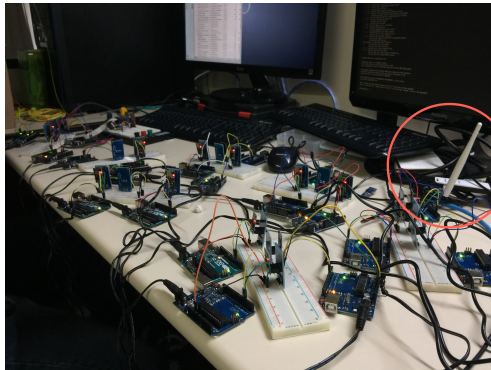


Figure 7. Testbed with 20 sensing nodes and partial view of the fog node (inside the red circle).

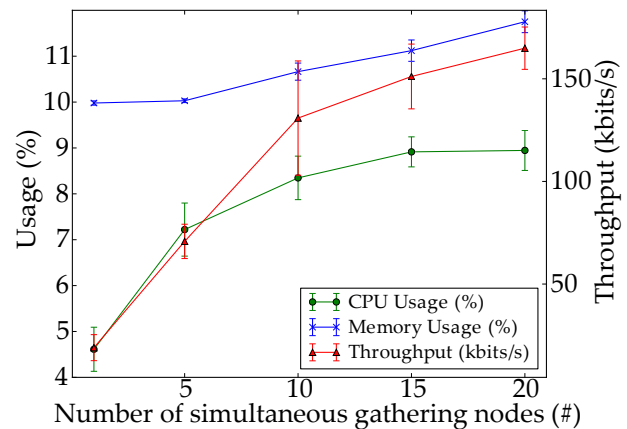


Figure 8. Resources usage and throughput in the fog node as a function of the number of devices.

4. **Cruz, P.**, Couto, R. S., and Costa, L. H. M. K. “An algorithm for sink positioning in bus-assisted smart city sensing.” In **Future Generation Computer Systems**, pp. 761-769, Oct. 2017. QCC: A2 ; QEIV: B1 IF: 6.125.
5. **Cruz, P.**, Silva, F. F., Pacheco, R. G., Couto, R. S., Velloso, P. B., Campista, M. E. M., and Costa, L. H. M. K. “SensingBus: Using Bus Lines and Fog Computing for Smart Sensing the City.” In **IEEE Cloud Computing**, pp. 58-69, Oct. 2018. CC: B2; QEIV: -; IF: 4.393.
6. **Cruz, P.**, Couto, R. S., and Costa, L. H. M. K. “Análise da Cobertura Espacial de uma Rede de Sensores Baseada em Ônibus Urbanos.” In **Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, May 2018. QCC: B2.
7. **Cruz, P.**, Couto, R. S., Lucena, A., and Costa, L. H. M. K. “Estratégias de pré-processamento para posicionamento de pontos de coleta em redes de sensores móveis.” In **Simpósio Brasileiro de Pesquisa Operacional**, Aug. 2018.
8. **Cruz, P.**, Couto, R. S., Costa, L. H. M. K., Fladenmuller, A., and Amorim, M. D. “On the Coverage of Bus-Based Mobile Sensing.” In **Sensors**, pp. 1-12, Jun. 2018. QCC: A1; QEIV: A1. IF: 3.031.
9. **Cruz, P.**, Couto, R. S., Costa, L. H. M. K., Fladenmuller, A., and Amorim, M. D. “A delay-aware coverage metric for bus-based sensor networks.” In **Computer Communications**, pp. 192–200, Jun. 2020. QCC: A2; QEIV: A1. IF: 2.816.
10. **Cruz, P.**, Couto, R. S., Costa, L. H. M. K., Fladenmuller, A., and Amorim, M. D. “Per-Vehicle Coverage in a Bus-Based General-Purpose Sensor Network.” In **IEEE Wireless Communications Letters**, Jun. 2020. QCC: B1; QEIV: A1. IF: 4.660.

In the context of this thesis, we have also coauthored two publications as a result of the contribution to the work led by other colleagues of the lab:

1. Couto, R. S., Barreto, H. F. S. S. M., **Cruz, P.**, Silva, F. F., Sciamarella, T., Campista, M. E. M., Costa, L. H. M. K., and Rubinstein, M. G. “Building an IaaS cloud with droplets: a collaborative experience with OpenStack.” In **Journal of Network and Computer Applications**, pp. 1-11, May 2017. CC: A2 ; QEIV: A1; IF: 5.570.
2. Ortiz, F. M., **Cruz, P.**, Couto, R. S., and Costa, L. H. M. K. “Caracterização de uma Rede Sem-fio de Baixa Potência e Longo Alcance para Internet das Coisas.” In **Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**, May 2018. QCC: B2.
3. Silva, F. D. de M., **Cruz, P.**, Couto, R. S., and Costa, L. H. M. K. “Redução de Inconsistências no Monitoramento da Frota de Ônibus Urbanos.” In **Simpósio Brasileiro de Telecomunicações e Processamento de Sinais**, Nov. 2020.

The software of the SensingBus prototype is registered with the INPI (Instituto Nacional da Propriedade Industrial - Brazilian Institute of Industrial Property), under the number BR512019002626-8.

References

- Ali, J. and Dyo, V. (2017). Coverage and mobile sensor placement for vehicles on predetermined routes: a greedy heuristic approach. In *ICETE 2017*. SCITEPRESS.
- Church, R. and Velle, C. R. (1974). The maximal covering location problem. *Papers in regional science*, 32(1):101–118.
- Cruz, P. (2020). *A bus-based opportunistic sensing network*. PhD thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro.
- Ekici, E., Gu, Y., and Bozdag, D. (2006). Mobility-based communication in wireless sensor networks. *IEEE Communications Magazine*, 44(7):56 – 62.
- Gao, Y., Dong, W., Guo, K., Liu, X., Chen, Y., Liu, X., Bu, J., and Chen, C. (2016). Mosaic: A low-cost mobile sensing system for urban air quality monitoring. In *INFOCOM'2016*, pages 1–9. IEEE.
- Ghafoor, S., Rehmani, M. H., Cho, S., and Park, S.-H. (2014). An efficient trajectory design for mobile sink in a wireless sensor network. *Computers & Electrical Engineering*, 40(7):2089–2100.
- Kariv, O. and Hakimi, S. L. (1979). An algorithmic approach to network location problems. i: The p-centers. *SIAM Journal on Applied Mathematics*, 37(3):513–538.
- Li, W., Santos, I., Delicato, F. C., Pires, P. F., Pirmez, L., Wei, W., Song, H., Zomaya, A., and Khan, S. (2016). System modelling and performance evaluation of a three-tier cloud of things. *Future Generation Computer Systems*.
- Liu, B., Brass, P., Dousse, O., Nain, P., and Towsley, D. (2005). Mobility improves coverage of sensor networks. In *MobiHoc 2005*. ACM.
- Marjovi, A., Arfire, A., and Martinoli, A. (2015). High resolution air pollution maps in urban environments using mobile sensor networks. In *DCOSS'2015*. IEEE.
- Sanchez, L., Muñoz, L., Galache, J. A., Sotres, P., Santana, J. R., Gutierrez, V., Ramdhany, R., Gluhak, A., Krco, S., Theodoridis, E., et al. (2014). SmartSantander: IoT experimentation over a smart city testbed. *Computer Networks*, 61:217–238.
- Umer, T., Amjad, M., Afzal, M. K., and Aslam, M. (2016). Hybrid rapid response routing approach for delay-sensitive data in hospital body area sensor network. In *ICC-CNT'2016*. ACM.
- Wong, J. L., Jafari, R., and Potkonjak, M. (2004). Gateway placement for latency and energy efficient data aggregation. In *LCN'2004*, pages 490–497. IEEE.
- Zanella, A., Bui, N., Castellani, A., Vangelista, L., and Zorzi, M. (2014). Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1):22–32.
- Zhao, D., Ma, H., Liu, L., and Zhao, J. (2013). On opportunistic coverage for urban sensing. In *MASS'2013*, pages 231–239. IEEE.
- Zoysa, K. D., Keppitiyagama, C., Seneviratne, G. P., and Shihan, W. W. A. T. (2007). A public transport system based sensor network for road surface condition monitoring. In *NSDR'2007*. ACM.