Offloading Floating Car Data

Razvan Stanica
Université de Lyon, INRIA UrbaNet
INSA-Lyon, CITI Lab
razvan.stanica@insa-lyon.fr

Marco Fiore
CNR-IEIIT and INRIA UrbaNet
marco.fiore@ieiit.cnr.fr

Francesco Malandrino
DET
Politecnico di Torino
malandrino@tlc.polito.it

***Abstract*—Floating Car Data (FCD) is currently collected by moving vehicles and uploaded to Internet-based processing centers through the cellular access infrastructure. As FCD is foreseen to rapidly become a pervasive technology, the present network paradigm risks not to scale well in the future, when a vast majority of automobiles will be constantly sensing their operation as well as the external environment and transmitting such information towards the Internet. In order to relieve the cellular network from the additional load that widespread FCD can induce, we study a local gathering and fusion paradigm, based on vehicle-to-vehicle (V2V) communication. We show how this approach can lead to significant gain, especially when and where the cellular network is stressed the most. Moreover, we propose several distributed schemes to FCD offloading based on the principle above that, despite their simplicity, are extremely efficient and can reduce the FCD capacity demand at the access network by up to 95%.**

I. Introduction

Floating Car Data (FCD) consist of information generated by moving vehicles and uploaded to Internet-based control centers for processing and analysis. FCD is today employed for, e.g., distant monitoring of on-board Electronic Control Units (ECUs). ECUs, whose number varies between 30 for low-end cars and 100 for premium-class automobiles [1], locally oversee in-vehicle operations, controlling almost all car functionalities. Systems such as BMW Assist, Ford SYNC, General Motor OnStar, Toyota Safety Connect and Mercedes-Benz mbrace, just to cite a few representative examples, retrieve the data generated by the ECUs in the form of FCD, so as to provide seamless distant support to the driver and the passengers. Services cover safety, diagnostic and anti-theft applications. Another common practical use of FCD is in the field of real-time road traffic monitoring. As an example, technologies such as TomTom HD Traffic and Meihui TrafficCast leverage FCD carrying anonymized vehicle position and speed so as to determine the traffic conditions in real-time and offer more efficient navigation services.

The FCD upload is today performed individually by each participating vehicle via the cellular infrastructure, as shown in Fig. 1a. Since the present negligible penetration rates of FCD-based technologies allow the cellular infrastructure to accommodate the FCD uplink traffic, completely relying on the pervasive access network is a convenient practice.

However, the success of the few FCD-based services that have been deployed is fostering a large-scale adoption of FCD-based solutions. As a significative example, FCD is going to play a key role in pervasive urban sensing: vehicles would collect environmental information about the metropolitan areas they travel through, and upload such data to Internet-based control centers for fusion and analysis. Urban sensing is indeed envisioned to significantly improve our understanding of urban dynamics and is commonly regarded as a fundamental component in forthcoming smart cities.

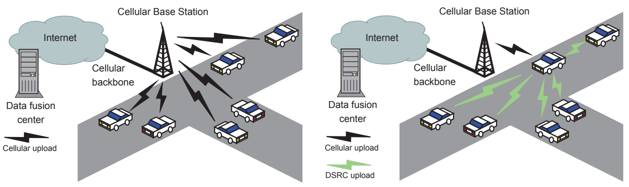
Additionally, many of the existing and future usages of FCD require the harvesting of data from the largest possible vehicle population. For example, for the distant monitoring of ECUs, each single car must be continuously probed, while in real-time traffic monitoring or in urban sensing, the quality of the aggregate information significantly improves with the number of available (positioning or sensing) samples. Not only the quantity, but also the frequency of the FCD collection is foreseen to progressively increase, driven by the need for growing accuracy in the monitoring or sensing activities.

These observations let us reasonably speculate that the consolidation of FCD-based technologies reaching near-100% penetration rates risks to induce a non-negligible load on the cellular uplink access. Considering that the mobile demand has reached the capacity limits of 3G networks [2], and that even the upcoming LTE infrastructure is already deemed unable to cover such a gap [3], offloading FCD could only benefit the cellular network operation.

To that end, direct vehicle-to-vehicle (V2V) communication based on Dedicated Short Range Communication (DSRC) could come in handy. The standardization activity has recently led to a number of protocol stack proposals, including IEEE 802.11p, IEEE 1609.x, ETSI ITS G5 and ISO CALM, deemed to enable communication among vehicles traveling within a range of a few hundred meters. Local gathering and fusion via V2V transfers could be leveraged for FCD offload as shown in Fig. 1b. There, a subset of the vehicles gather the data sensed by neighboring cars through DSRC communication, and fusion it with their own observations before uploading the aggregate information via a single cellular network transfer. The system would result in significant offload of the cellular infrastructure, in terms of channel signalization, control header overhead and, depending on the local fusion level, sheer uplink traffic volume.

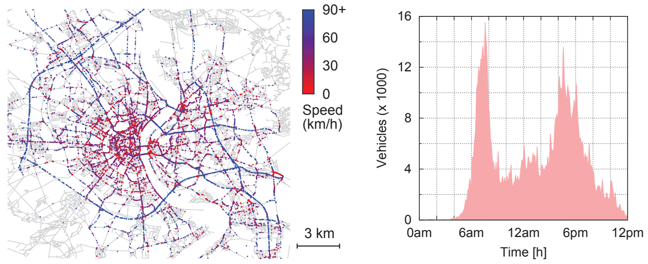
In this paper, we explore the above a scenario for offloading the cellular infrastructure from FCD uploads. This is, to the best of our knowledge, the first work to address this problem. FCD has been considered as a research topic mainly in transportation and traffic planning theory [4]. From a networking perspective, it has been tackled in terms of user privacy [5], however the interaction of FCD and network infrastructure has never been considered before.

The previous literature on cellular network offloading has



(a) Traditional FCD upload (b) Offloaded FCD upload

Fig. 1. FCD upload scenarios: traditional and offloaded through V2V transfer.



(a) Road traffic snapshot (b) Road traffic volume

Fig. 2. Traffic characteristics for the vehicular trace from the city of Köln

instead focused on downlink data transfers, targeting the dissemination of content to smartphone users [6] or car passengers [7] and the download of large contents by vehicular users [8]. Their downlink nature make these problems semantically different from ours; moreover, all of the works above consider delay-tolerant approaches, while FCD usages, such as those we outlined in Sec. I, typically require the upload to occur in quasi real-time.

Other works have addressed the efficiency of offloading part of the cellular traffic through WiFi networks [9]. However, the use of WiFi access points by passing vehicles is by its own nature an opportunistic solution, that can help for example in the case of in-vehicle web access, but it is not reliable enough for a number of applications that need to collect FCD with low delay and fine granularity in both time and space.

The FCD offload problem is thus different from those of downlink or WiFi-based offloading, requiring a dedicated analysis. More precisely, offloading FCD uploads maps to (i) *identifying the subset of vehicles in charge of performing the data fusion and upload, so as to harvest the maximum FCD amount*, and (ii) *doing so in an efficient distributed way*.

To solve these problems, we proceed as follows. We first present the large-scale vehicular scenario used in our study, describe it through a time-varying graph model and observe several fundamental properties of V2V connectivity that we later exploit to study the FCD offloading problem, in Sec. II. We then approach the problem from an oracle viewpoint, formulate FCD offloading as an optimization problem and derive the optimal solution, in Sec. III. Several practical distributed schemes are proposed and compared to the optimal in Sec. IV. Finally, conclusions are drawn in Sec. V.

II. Scenario, model and connectivity

Our study leverages a realistic vehicular mobility dataset of a large-scale urban region, presented in Sec. II-A. Direct communication among vehicles allows us to model the road traffic as a time-varying connectivity graph as detailed in Sec. II-B. The fundamental properties of such a graph that are of interest to our analysis are discussed in Sec. II-C

A. Road traffic scenario

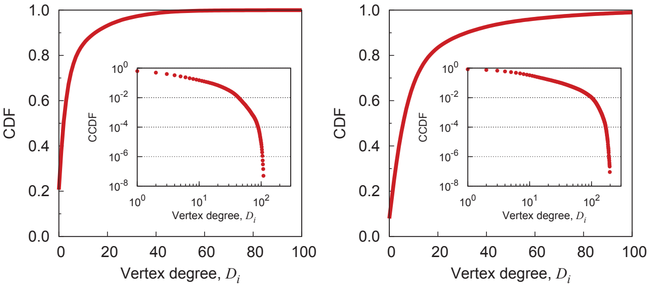
The mobility dataset we employ reproduces the road traffic in the greater urban area of Köln, in Germany. It covers 4500 km of roads in an area of 400 km^2, and spans over 24 hours of a typical working day. As a result, it includes information on more than 700.000 car trips, with per-second data on the position and speed of each vehicle. This is the largest and most complete vehicular mobility dataset freely available to date. A picture of the road topology, including a snapshot of the vehicular mobility at 7 am is provided in Fig. 2a, where each dot represents one car, its color corresponding to its speed. Fig. 2b portrays instead the daily evolution of the road traffic volume, expressed in thousands of vehicles.

The dataset is generated by coupling different state-of-art tools for the road topology information, microscopic mobility modeling and macroscopic traffic flow definition. Namely, the road topology data is obtained from the OpenStreetMap (OSM) database, commonly regarded as the highest-quality map database publicly available. The microscopic mobility of vehicles is simulated with the Simulation of Urban Mobility (SUMO) software, today's most advanced freely available microscopic vehicular mobility generator. Finally, the macroscopic traffic flows are determined in two steps. First, the Travel and Activity PAtterns Simulation (TAPAS) methodology [10] is applied on real-world data collected in the Köln region [11] to obtain a travel demand (i.e., the origin, destination and time of trips) that faithfully mimics the daily activity of the area residents. Then, Gawron's relaxation algorithm [12] is run to determine a traffic assignment (i.e., the routes taken by each driver) that allows a so-called dynamic user equilibrium. For further details on the dataset, we refer the reader to [13].

B. V2V connectivity graph model

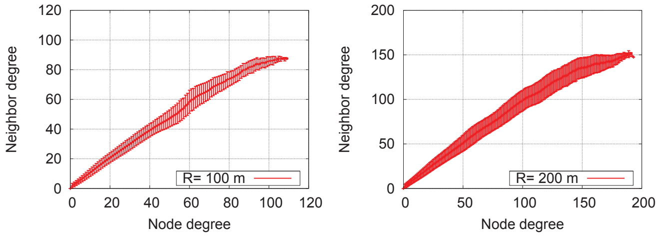
DSRC-based technologies allow vehicles to establish communication links among them, that can be used, as previously stated, to transfer FCD, other than data of different nature. For example, vehicles using DSRC natively obtain information regarding their neighbors by the means of safety beacons (standardized as ETSI CAM messages and by the SAE J2735 dictionary set) periodically transmitted on the so called *control channel*, dedicated to safety applications. Although some standards (e.g. IEEE 802.11p) claim that DSRC communication allows to reach distances up to 1 km, the extensive experimental analysis in [14] shows that, in an urban environment and with common power levels and modulation, vehicles at a distance of 100 m can exchange data with a 80% Packet Delivery Ratio (PDR), while the PDR drops to 50% when the distance increases to 200 m.

The way we choose to represent V2V connectivity is through a time-varying graph $G(\mathbb{V}(t), \mathbb{E}(t))$, where $\mathbb{V}(t) =$



(a) R = 100 m (b) R = 200 m

Fig. 3. Vertex degree distributions observed in the V2V connectivity graph of the Köln urban region, for different communication ranges.



(a) R = 100 m (b) R = 200 m

Fig. 4. The mean degree of a neighbor (with standard deviation) as a function of the node degree for the entire 24 hours of the trace.

$\{v_i\}$ is a set of vertices (also referred to as nodes in the following) v_i, each mapping to a vehicle i traveling in the scenario at time t, and $\mathbb{E}(t) = \{e_{ij}(t) \mid v_i, v_j \in \mathbb{V},\ i \neq j\}$ is the set of edges $e_{ij}(t)$, each representing a communication link established between vehicle i and vehicle j at time t.

In this work we will assume a simple unit disk model, adding an edge in $\mathbb{E}(t)$ whenever the distance between the two nodes is below a threshold R. We acknowledge that the unit disk graph is a drastic simplification of the reality when edges need to model node connectivity during short time periods, as signal propagation does not follow these simple rules in reality. However, we do not start from the hypothesis that $G(\mathbb{V}(t), \mathbb{E}(t))$ represents the instantaneous connectivity of the vehicular network at time t. In our case, an edge between v_i and v_j simply means that the communication link between i and j is sufficiently reliable to allow the two vehicles to have a good estimation of each other's position. This is usually the case even for relatively small PDR values, implying that an edge can exist even when the corresponding radio link is temporarily down. We consider a PDR of 50% is a reasonable lower limit for the quality of a link in this model, therefore we perform our study considering two values for R, namely 100 and 200 meters. As vehicles implicitly know the position of other vehicles connected through a good link, they can simply filter those situated at distances higher than R, so the graph G is very easy to construct. To summarize, the only assumption we make in our model is that a vehicle i has sufficiently reliable links with all the vehicles situated closer than R.

C. Fundamental connectivity properties

In our study, we will leverage two measures derived from the V2V connectivity graph. As we are interested in the local gathering and fusion of FCD, both measures concern the communication neighborhood of a vehicle i.

The first measure is the *vertex degree*, that relates to the one-hop neighborhood of a generic vehicle i. Formally, let us consider a subset $\mathbb{V}_i^1(t) = \{v_j \mid \exists\, e_{ij}(t)\}$ of vertices and a subset $\mathbb{E}_i^1(t) = \{e_{jk}(t) \mid v_j, v_k \in \mathbb{V}_i^1(t)\}$ of edges. We define the subgraph $C_i^1(t) = G(\mathbb{V}_i^1(t), \mathbb{E}_i^1(t))$ as the one-hop communication neighborhood of vertex i at time t. The degree of a vertex v_i is then $D_i(t) = \left\|\mathbb{V}_i^1(t)\right\|$, where $\|\cdot\|$ denotes the cardinality of the included set.

The probability distributions of the vertex degree measured in the V2V connectivity graphs extracted from the Köln scenario are depicted in Fig. 3. The outset plots portray the Cumulative Distribution Function (CDF) of the vertex degree. We can observe that a significant proportion of the nodes have a relatively low number of communication neighbors: when R = 100 m, 90% of vertices have a degree less than 10, while with R = 200 m, such a percentage is at 75%. The inset plots show instead the Complementary CDF (CCDF) in a log-log scale. Those evidence the appearance, although with low probability, of nodes with high degrees, up to 120 when R = 100 m and up to 170 when R = 200 m.

The second measure we are interested in is the *vertex assortativity*, and concerns the two-hop neighborhood of vehicle i. Formally, the assortativity is defined as $A_i(t) = \frac{1}{D_i(t)} \sum_{j \in \mathbb{V}_i^1(t)} D_j(t)$. In other words, the assortativity of a vehicle is the average vertex degree of its one-hop communication neighbors. It is especially important to observe the relationship between $D_i(t)$ and $A_i(t)$. Plots correlating these two measures for all the samples (i.e., i and t pairs) observed over the 24h are portrayed in Fig. 4, for both R = 100 m and R = 200 m. The average behavior, pointed out by the solid line, evidences the strong linear correlation between $D_i(t)$ and $A_i(t)$. In complex network theory, this phenomenon is referred to as *network assortativity* [15], and implies that high-degree nodes are connected to similar high-degree nodes, while low-degree vertices are mainly connected to other low-degree vertices.

In the following, we will separately study the network at each time instant. Thus, for the sake of clarity, we will drop the dependence from time t in the notation and refer to the graph at the generic current time instant.

III. Optimal Gain

Since the FCD offloading problem was not previously addressed in the literature, we are first interested in understanding how much one can hope to gain when switching from the traditional cellular-based operation to the V2V-based one. To that end, we formulate FCD offloading as an optimization problem and derive the optimal system performance.

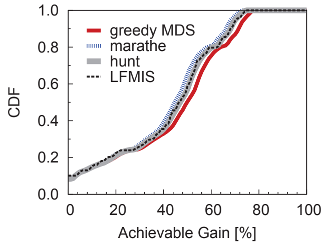
Firstly, we recall that our goal is to identify a set of vehicles, each of which gathers FCD from its communicating neighbors, and performs the data fusion and upload. Clearly, we wish such a set to be as small as possible, since the fewer the vehicles performing the local fusion and upload, the lower the uplink traffic load on the cellular network. At the same time, however, we do not want the offload process to reduce the quality of the overall FCD information: this maps to the requirement that the V2V gathering of Fig. 1b should collect the same FCD that would have been individually uploaded by cars in the infrastructure-based approach of Fig. 1a.

In other words, the objective becomes that of collecting FCD from the whole network using the least amount of vehicles, so as to minimize the number of uploads and maximize the local FCD fusion. The FCD offloading problem can thus be formulated as a *Minimum Dominating Set* problem, whose solution yields the minimal set of vehicles that cover all the other cars through V2V communication,

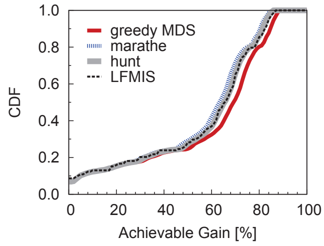
Formally, given the vehicular network graph $G(\mathbb{V}, \mathbb{E})$ at a generic time instant, a *Dominating Set* $\mathbb{S} \subseteq \mathbb{V}$ is defined as $\mathbb{S} = \{v_i \mid \exists\ e_{ij}\ \forall v_j \in \mathbb{V} \setminus \mathbb{S}\}$ i.e., each vertex in $\mathbb{V} \setminus \mathbb{S}$ has at least an edge towards a vertex in $\mathbb{S}$. If removing any vertex from $\mathbb{S}$ breaks this dominance property, then $\mathbb{S}$ is a *Minimal Dominating Set*. Finally, the *Minimum Dominating Set* (MDS), is the Minimal Dominating Set with the smallest size and represents the optimal solution to our coverage problem.

On both general and unit disk graphs, the problem of finding a MDS is NP-hard. However, a very simple greedy algorithm exists, able to compute a lnD_{max}-approximation of the MDS, where D_{max} denotes the maximal degree of graph $G(\mathbb{V}, \mathbb{E})$ [18]. A simplified version of the greedy algorithm, known as the Lexicographically First MIS (LFMIS) [17], produces an MDS by using the related problem of the *Maximum Independent Set* (MIS). Other heuristics, like the technique proposed by Marathe *et al.* [19], or the Polynomial-Time Approximation Scheme (PTAS) described by Hunt *et al.* [20] leverage properties of unit disk graphs to reduce the approximation with respect to the optimal MDS solution.

In the following, we will evaluate different solutions to the MDS problem in terms of *system gain*, i.e., *the fraction of vehicles that do not have to access the cellular infrastructure when FCD is offloaded through DSRC communication*. Such a metric has the significant advantage of being very intuitive and of general validity throughout different cellular technologies and FCD applications. Indeed, considering a specific infrastructure (e.g., GPRS, UMTS or LTE) impacts the cost of the channel setup per FCD upload, and thus maps to a simple scaling factor to our gain metric. Diverse FCD applications allow instead for different levels of local FCD compression at the gathering vehicle. Let us take as examples the city-wide estimation of the pollution level and the collection of exact car positions in a precise area. Clearly, the first use case allows the collecting vehicle to reduce its neighbors' observations to one average value and thus to significantly cut the traffic volume uploaded to the cellular network. Conversely, the second application requires in all cases the transmission of the precise location of each car, thus reducing the local fusion advantage to using a single header stack for the whole set of collected position information. Therefore, also the application can be mapped to a scaling factor to our gain metric.



(a) R = 100 m



(b) R = 200 m

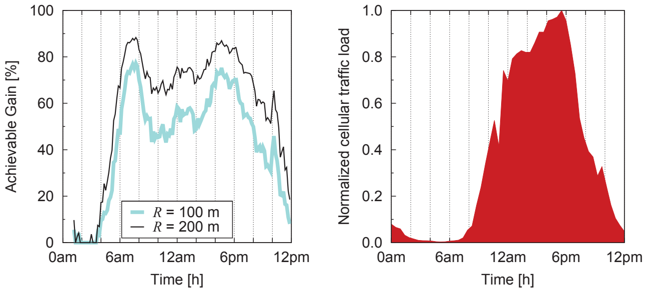
Fig. 5. CDF of the achievable gain during 24 hours of vehicular mobility for the two V2V communication ranges.

Overall gain. In order to estimate the gain that can be achieved by an FCD offload mechanism, we apply the four MDS algorithms mentioned above to our vehicular trace. Fig. 5 shows the distribution of the gain, or, more precisely, the Cumulative Distribution Function (CDF) of the non-uploading cars over the whole day, for the different MDS heuristics. Two major aspects should be pointed out in this figure. First of all, although the greedy MDS obtains the sets with the smallest size, all the four heuristics, with different approximation capabilities in theory, lead to relatively close results, indicating that the obtained relay sets are close to the optimal solution. Second, the overall gain can be significant: in more than 70% of the cases we are theoretically able to offload more than 60% of the FCD when a V2V communication range of 200 m is used. In some situations (discussed below) the gain can even reach more than 90%.

Impact of daytime. The results above are unrolled over daytime in Fig. 6a. We can note that the gain varies significantly depending on the hour considered: the gain is lower when the road traffic activity volume, in Fig. 2b, is lower, and it grows as the presence of vehicles in the Köln region increases. As one can expect, when the number of cars is very low, e.g., before 5 am or after 11 pm, most vehicles are isolated from a V2V communication viewpoint, and thus forced to upload their own FCD individually. However, this is not a major problem, as the number of implicated vehicles is extremely low, and the cellular infrastructure is almost not utilized during those time periods, as also proven by Fig. 6b, portraying the typical daily load pattern of a cellular base station, as measured in real-world urban cells from an anonymous operator.

More interestingly, V2V communication can offload from 40% to 70% of FCD during non-rush traffic hours, i.e., in the time intervals 9 am – 3.30 pm, and 7 pm to midnight. The former time period in particular matches to moderate-to-high data traffic load according to Fig. 6b, and thus FCD offload could have a significant impact there.

Finally, it is during the traffic peak periods, i.e., between 6.30 am and 9 am and between 3.30 pm and 7 pm, that the DSRC-based FCD offloading attains its maximum gain, around 75% when R = 100 m and up to 90% when R = 200 m. While the first period is a low-load one for the cellular infrastructure, the second maps exactly to the daily peak in the cellular data traffic load. We can thus conclude that during several hours in the afternoon high-gain FCD offload may



(a) FCD offload

(b) Cellular traffic load

Fig. 6. Time dynamics of FCD offload: (a) optimal gain with different R's; (b) typical normalized traffic load at the cellular access network, over a 24-hour period. The latter data was obtained from experiments conducted by the Autonomous Networks Research Group at USC, `http://anrg.usc.edu`.

prove paramount to reduce the load on the infrastructure.

Impact of geographical areas. We analyze the impact of FCD offloading from a geographical perspective in Fig. 7, where the plots portray the gain in each of the 86 Köln districts at key day hours, as obtained with the greedy MDS selection of uploading vehicles.

It is evident that, no matter the time considered, the distribution of the gain is far from being uniform over the region. The Köln city center is the area with the highest gain, reaching values around 0.95 at 7 am, which means that 95% of the vehicles can avoid accessing the cellular network and still contribute to the FCD collection. The South-East suburban area, where Porz, the largest borough of Köln, is situated, is also characterized by consistently higher gain than most other areas. Other districts, such as the southern ones, show instead low FCD offloading gain over the whole day.

Once more, the gain is largely dependent on the road traffic volume. The downtown or Porz areas are where most of the vehicular activity gathers, which results in higher vertex degrees of the V2V connectivity graph and higher margin of gain through the MDS uploader selection. Most importantly, these regions are again those characterized by the highest human activity, and thus by the highest cellular network load. Therefore, these are also where the FCD offloading will be more critical.

Overall, our results showed that FCD offloading through V2V communication is an interesting approach to reduce the volume of traffic that will be uploaded to the cellular infrastructure by sensing vehicles in urban areas, and this DSRC-based FCD offloading performs best precisely at the hours and in the regions where the cellular network needs to be relieved the most.

IV. Practical solutions

Having identified the significant gain attainable through DSRC-based FCD offloading, we are now interested in defining practical solutions that exploit such a networking paradigm. This implies relaxing the oracle assumption and designing distributed solutions to the MDS problem in a dynamic vehicular network.

Research on distributed local algorithms for the MDS problem has been flourishing in the last decade, mostly thanks to its possible applications in wireless sensor networks. Usually, these solutions manage to find a constant approximation of the MDS, but require a communication cost that depends on the size of the network. The trade-off between the quality of the approximation and the amount of needed communication in the case of unit disk graphs has been investigated by Kuhn *et al.* [16], where one of the best distributed algorithms for the dominating set problem is also proposed. Their solution leads to an approximation factor of $O(D_{max}^{1/\sqrt{cr}} log D_{max})$ using $O(cr)$ communication rounds.

However, the concept of *constant time-approximation* (i.e. constant number of communication rounds) needs to be understood in the context of the synchronous message passing model used in all these studies coming from the distributed systems community. This model makes the assumption that, in every round, each node can send a different message to each of its neighbors. While an algorithm requiring a multiple (and constant) number of rounds can be interesting in a static scenario like the initialization of a wireless sensor network, such a solution does not seem suited for networks involving high mobility, where the network topology would change before the algorithm reaches a solution.

As opposed to these general MDS algorithms, in this section we propose and analyze three heuristics that allow the distributed construction of a set of relays used to offload FCD in a time frame that takes into account the properties of direct V2V communications.

A. Degree-Based (DB)

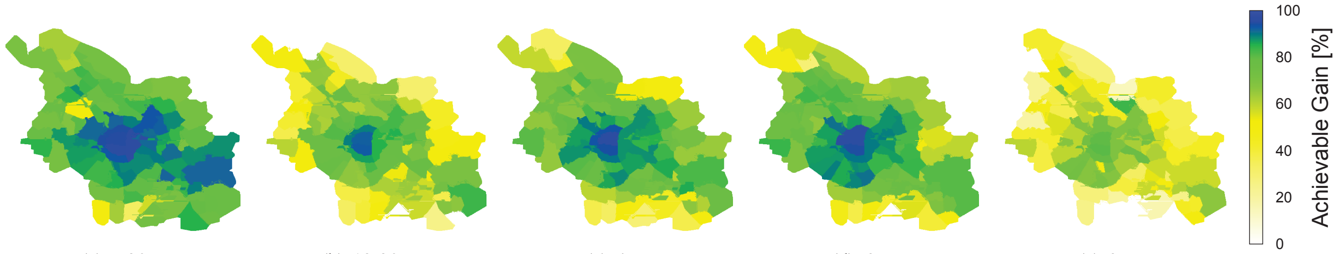
The first mechanism we describe uses the safety beacons received on the control channel to compute the degree of the nodes in the vehicular network and then decides that every vehicle belongs to the transmitting set with a probability that depends on the number of neighbors it possesses. This means that, if we consider a node v_i with D_i neighbors (including vehicle i itself, therefore $D_i \geq 1$), i transmits on the cellular link with a probability k/D_i, where k is an important parameter for the trade-off between coverage and offloading gain (the optimal value for k is discussed below). While the mechanism does not build a Dominating Set and therefore can not provide any guarantee on the coverage of the entire area, its simplicity makes the study of its performance very intriguing.

Under the assumptions presented above, the probability that a node v_i with $D_i = d$ is not covered by any transmission depends on $P_{nt}(d)$, the probability that the node itself does not transmit, and $P_{nn}(d)$, the probability that all its one hop neighbors decide against using the cellular uplink:

$$P_{nc}(d) = P_{nt}(d) \cdot P_{nn}(d). \tag{1}$$

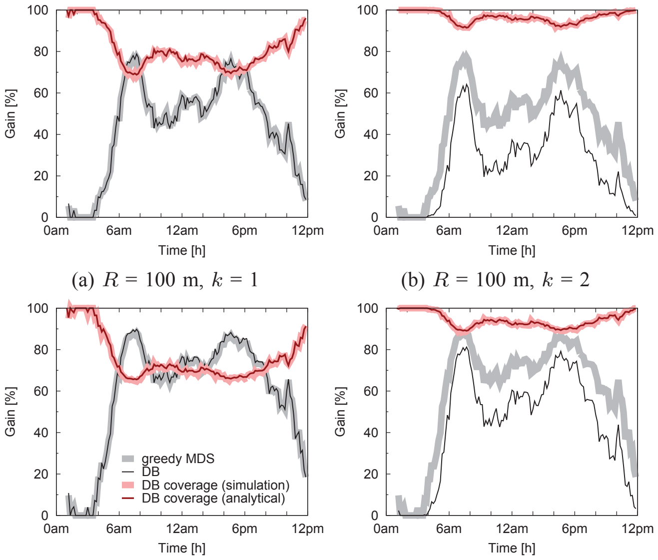
The probability that vehicle i does not transmit in this scenario can be written as follows:

$$P_{nt}(d) = 1 - min\,(1, k/d)\,.$$



(a) 7:30am (b) 12:30am (c) 4pm (d) 6pm (e) 9pm

Fig. 7. Spatial dynamics of FCD offload: the FCD offloading gain granted in each of the 86 *Stadtteile*, i.e., districts, of Köln at key hours of the day, when the greedy MDS heuristic is employed for the selection of the uploading vehicles (best viewed in color).



(a) R = 100 m, k = 1 (b) R = 100 m, k = 2

(c) R = 200 m, k = 1 (d) R = 200 m, k = 2

Fig. 8. The gain obtained by using the DB mechanism compared to the situation with no offloading, and the ratio of vehicles covered by the mechanism, for different values of the vehicular transmission range and of the transmission probability. The legend in 8c applies to all the plots.

Considering $\mathbb{V}_i^1$ to be the set containing all the one-hop neighbors of node v_i, the last term of Eq. 1 becomes:

$$P_{nn}(d) = \prod_{j:v_j \in \mathbb{V}_i^1} (1 - min\,(1, k/D_j)).$$

However, as discussed in Section II-C, vehicular networks are assortative, with links between nodes of similar degree. Therefore we can consider that $D_j \approx D_i = d$ and, by replacing the terms in Eq. 1, we obtain:

$$P_{nc}(d) \approx (1 - min\,(1, k/d))^d\,. \tag{2}$$

This means that the ratio of nodes not covered by this transmission scheme depends on k and can be written as:

$$r_u(k) = \sum_{d=1}^{\infty}(\pi_d \cdot P_{nc}(d)) \approx \sum_{d=k+1}^{\infty} \pi_d \cdot \left(1 - \frac{k}{d}\right)^d \tag{3}$$

where π_d represents the probability that a node has a degree d, and we took into account the fact that the terms with $k > d$ do not contribute to the sum.

Fig. 8 shows the performance of the proposed DB mechanism for the vehicular trace from the Köln region considering two different transmission probabilities ($1/D_i$ and $2/D_i$) and for two different coverage areas (100 m and 200 m). For every second of the trace, we build the corresponding graph, we compute the MDS using the greedy algorithm as a reference value, we calculate the node degree distribution and use it in Eq. 3 to obtain the analytical results. Finally, we repeat the DB selection algorithm 100 times to calculate the mean number of transmitters and the mean number of covered nodes.

In this figure, we can observe that the gain achieved by the DB mechanism when k=1 matches very well the one obtained by the greedy MDS. However, we must recall that the greedy MDS algorithm achieves 100% coverage, while the DB approach only covers around 70% of the vehicles with k=1. When we use a higher cellular transmission probability, e.g. k=2, the achieved gain is smaller, but more than 90% of the map is covered. Also, it is important to notice that the analytical results regarding the coverage achieved by the DB mechanism are practically identical with those obtained through simulation, allowing us to determine immediately the expected coverage of this solution.

B. Degree-Based with Confirmation (DB-C)

As shown by the results in the previous section, a probabilistic approach like the one based on a node's degree can not achieve a coverage of 100%. However, such a property might be required by some applications, and the DB approach can be extended by a simple confirmation mechanism in order to obtain this total coverage: a vehicle choosing to act as a relay for its neighbors also transmits a message on the V2V channel, announcing this to its neighbors. Therefore, during each data collection period, every node learns if it is covered by a neighbor or not. If the latter is true, the vehicle can transmit its own information on the cellular uplink, just as in the case when no V2V communication is used. The problem in this case is to distinguish the optimal value of the uplink transmission probability, and therefore in the following we transform the DB-C mechanism into an optimization problem.

In the proposed scenario, the two parameters of the optimization problem are the nodes selected to use the cellular uplink, and the network topology. We consider that the traditional upload scenario implies a constant cost of C_c, while acting as a relay for d neighbors leads to a cost of $C_v(d)$. As we are interested in the cellular uplink usage, a zero cost for V2V communication is assumed. In this case, we can compute an average *transmission cost*, over both varying parameters (selected nodes and topology), for each FCD collection period:

$$C = \sum_{d=1}^{\infty} \pi_d \left[C_c \cdot P_{nc}(d) + C_v(d) \cdot (1 - P_{nc}(d)) \right]. \quad (4)$$

Without loss of generality, we can assume that the cost in each scenario is given by the number of transmissions on the uplink, namely $C_c = 1$, and $C_v(d) = min(1, k/d)$. Using the value given in Eq. 2 for $P_{nc}(d)$, we can distinguish two different situations, with the node degree being either below or above the parameter k that gives the transmission probability on the cellular uplink. We write Eq. 4 to better represent these cases:

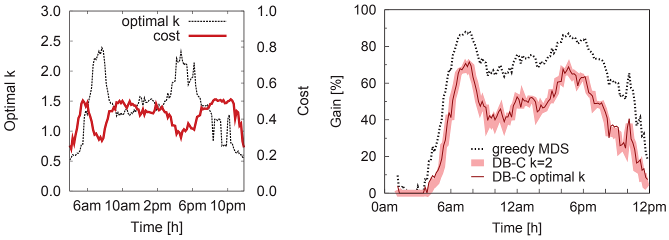
$$C = \sum_{d=1}^{k} \pi_d + \sum_{d=k+1}^{\infty} \left[\pi_d \frac{k}{d} + \pi_d \left(1 - \frac{k}{d}\right)^{d+1} \right]. \quad (5)$$

In Eq. 5, the first sum is the cost brought by those nodes which are sure to be selected by the DB mechanism because they have k-1 or less neighbors. The second term, of the second sum is the cost of those transmitting on the cellular uplink using the DB-C approach, as they remained uncovered by DB. Finally, the first term of the second sum gives us a gain over the traditional scenario, and represents the cost of those nodes relaying FCD for their neighbors in the first transmission round

It is not trivial to minimize the expression in Eq. 5 efficiently, as the variable k appears an unbounded (a priori) number of times, with different signs and exponents. The first fact, i.e., the $(1 - \frac{k}{d})^{d+1}$ term, makes it quite complex (although not entirely impossible) to analytically derive the expression and determine its minimum. The second fact, i.e., that k happens to appear with a negative coefficient, rules out other straightforward optimization techniques such as convex optimization of posynomial functions [21].

On the other hand, the expression in (5) has three desirable properties: (i) it is univariate, i.e., includes a single variable k; (ii) it is continuous and derivable with respect to k; (iii) the values of k can be bounded, e.g., between 0 and D_{max}. Root-finding and optimization of univariate expressions has been long studied, and there are plenty of existing solutions to the problem – most of which are similar, in principle, to the well-known bisection method. "Well-behaved" functions, especially derivable ones, are associated to faster convergence times. Finally, if the optimum can be searched for in a limited interval, the result can be guaranteed to be a global optimum (as opposed to a local one). In our case, we minimize Eq. 5 via the Brent method [22, Chapter 7], which runs in polynomial (namely, quadratic) time and returns a global optimum.

The optimization results are presented in Fig. 9 for a V2V communication range of 200 m (results for R= 100 m are consistent, and are not included because of space restrictions). First of all, Fig. 9a shows that the optimal value of the transmission probability is obtained for a value k generally between 1.5 and 2.5, and it follows the evolution of the vehicular density, with two peaks at around 7am and 6pm. On the other hand, the cost achieved by this optimal k is smaller when the number of vehicles increases, as more information is offloaded from the uplink.



(a) Optimal k and minimal cost (b) Selected transmitters

Fig. 9. Performance of the confirmation mechanism for R = 200 m.

Second, Fig. 9b compares the number of transmitters on the cellular network chosen by the DB-C (for an optimal k and for k=2) with the one obtained using the greedy MDS. We can notice that DB-C with k=2 produces results very close to the ones obtained for an optimal k, which is important from a practical point of view, as it allows us to use a single calibrated value for the transmission probability during the entire day. As discussed, DB-C is also able to cover the entire map, while reaching a gain of more than 60% at peak hours. However, these high density scenarios are also those where the mechanism has the most difficulties in approaching the optimal solution, with a difference of around 20%. Despite this issue, DB-C remains a very simple mechanism capable of offloading an important percentage of the FCD while covering all the vehicles in the region.

C. Reservation-Based (RB)

Although DB-C results in a 100% coverage, its performance can still be improved when the vehicular density increases. In order to reduce this gap, we propose a simple reservation mechanism which selects the relays in a single V2V communication round. As discussed below, this mechanism results in a constant approximation of the MDS in an ideal scenario and in small (although not minimal) Dominating Sets in a realistic DSRC-based network.

The proposed RB mechanism has the following steps:

- At the beginning of every collection period, there is a reservation phase (assumed slotted and containing N_s slots), where each vehicle selects a transmission slot among the N_s available and enters the *contender* state where it backs-off, waiting for the chosen slot.
- In every time slot, the vehicles in the *contender* state having chosen the corresponding back-off transmit a reservation message and change their status to *dominator*.
- A vehicle in the *contender* state and receiving a reservation message from one of its neighbors becomes *dominated* and cancels its back-off.

In the following, we study the performance of RB and we show that it manages to obtain a Dominating Set whose quality depends on the number of slots in the reservation period. As a matter of fact, in the ideal case, every node would choose a different transmission slot. If this condition is met, it is easy to see that the reservation mechanism is in fact a distributed implementation of the LFMIS algorithm. Every transmission of a *contender* is equivalent to the random selection of a node and, just like in the centralized version, all the one-hop neighbors become *dominated*. Therefore, as for the LFMIS algorithm, the obtained result is a constant approximation of the optimal MDS.

However, in a real distributed implementation it is extremely difficult to ensure that every node transmits in a different slot. Several solutions exist (e.g. [17]), but they all achieve the desired result by introducing several rounds of communication, an overhead that can not be sustained in a highly mobile vehicular network. On the other hand, the proposed RB mechanism is fast, but its result is no longer an MDS when synchronized transmissions are considered. However, in the following we show that the mechanism's performance depends on the number of slots in the reservation period and results close to those of LFMIS are obtained in realistic scenarios.

Before presenting the details of the model, it is important to understand the consequences of *synchronized transmissions*, the event of having multiple vehicles broadcasting reservation messages in the same slot. First of all, if two or more neighbors transmit in the same slot, RB can no longer guarantee that the obtained relay set is an MDS, as the selection of some nodes in the *dominator* set is simply an artifact of messages transmitted in parallel. Second, synchronized transmissions can also have an impact on the receivers, who might not be able to decode any of the reservation messages using the same slot, hence the impact of collisions must be considered.

In this case, let us consider the scenario of a reservation period with N_s slots, and a vehicle with d neighbors (itself included). Taking into consideration the assortativity property of vehicular networks, these neighbors also have a degree d. If $P_t(d,s)$ is the probability that a node with d neighbors transmits in slot s, we can calculate the ratio of vehicles using the cellular uplink as:

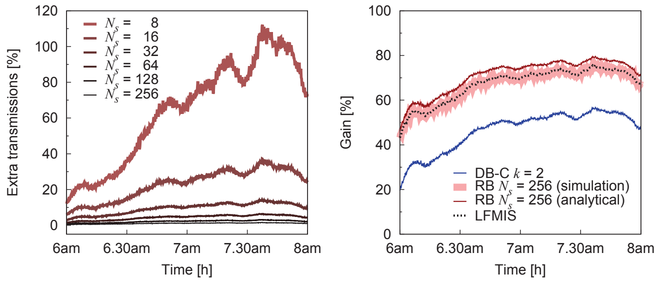
$$r_u = \sum_{d=1}^{\infty} \left(\pi_d \cdot \sum_{s=0}^{N_s-1} P_t(d,s) \right). \tag{6}$$

We denote as *covered* a node that is either a *dominator* or *dominated*, and $P_c(d,s)$ is the probability that a node of degree d becomes covered during slot s. Using a similar definition, $\bar{P}_c(d,s) = 1 - \sum_{i=0}^{s-1} P_c(d,i)$ represents the probability that a node has not been covered in the first $s-1$ slots.

In the first slot of the reservation period, the transmission probability only depends on the choice of the corresponding slot, therefore we have $P_t(d,0) = 1/N_s$. On the other hand, for any $s > 0$, the transmission probability depends on whether the node has already been covered in the previous slots and can be written as:

$$P_t(d,s) = \bar{P}_c(d,s) \cdot P_{tn}(d,s), \tag{7}$$

where $P_{tn}(d,s) = 1/(N_s - s)$ is the conditional probability of a transmission in slot s given that the node is not yet covered.



(a) Extra transmitters. (b) Achieved gain.

Fig. 10. Performance of the RB mechanism in the vehicular network scenario when the impact of collisions is considered. Figure 10a shows the percentage of extra transmitters when compared with the centralized LFMIS algorithm. Figure 10b presents the achievable performance when a reservation period of 256 slots is used. A V2V transmission range of 100 m is considered.

The last step in solving Eq. 6 consists in providing a formula for $P_c(d,s)$. However, this is a difficult task, even in an assortative network, as it would require knowing not only the distribution of the neighbors' degree, but also the distribution of the number of common neighbors between any adjacent nodes. Therefore, in order to keep our analysis tractable, we add a supplementary assumption, requiring the nodes to be grouped in cliques of degree d. In this classical example of an assortative network, there are two possibilities for a node to become covered. First of all, the node can transmit a reservation message (with probability $P_t(d,s)$). In this case, regardless the actions of the neighbors, the node becomes a dominator. The second option is that a single neighbor decides to broadcast its reservation during a given slot, avoiding any collision with other nodes. This second event happens with a probability $P_{sn}(d,s)$ and can be calculated as follows:

$$P_{sn}(d,s) = \bar{P}_c(d,s) \cdot \binom{d-1}{1} \cdot P_{tn}(d,s) \cdot (1 - P_{tn}(d,s))^{d-1}. \tag{8}$$

Using Eq. 8, the probability that a node becomes covered during slot s can be written as $P_c(d,s) = P_{sn}(d,s) + P_t(d,s)$.

An important property stems from the fact that $P_t(d,N_s) = \bar{P}_c(d,N_s)$, meaning that at the end of the reservation period, all the nodes will be covered. This proves that the proposed reservation mechanism results in a Dominating Set. However, in the extreme case when $N_s = 1$, the transmission probability is 1 for every vehicle, leading to a 100% ratio of users on the cellular uplink. Conversely, when N_s tends to infinity, the reservation mechanism is equivalent to the LFMIS algorithm.

In order to study the impact of the number of slots on the quality of the obtained solution, we simulate the reservation mechanism on 2 hours of the vehicular trace, between 6 am and 8 am (the moment when the performance of DB-C drops). Fig. 10 shows the performance of the RB mechanism in this scenario. The results on the left present the number of supplementary relays introduced by RB due to synchronized transmissions and collisions, when compared with the centralized LFMIS algorithm. We can notice that, when a small number of slots is used (e.g. $N_s = 8$), the difference is important, and the number of transmitters can even be doubled under high density. However, as the number of reservation slots increases, the number of extra-relays becomes less significant with a period of 256 slots resulting in the same results as the LFMIS algorithm. This close approximation of the LFMIS with a 256-slots period is also confirmed in Fig. 10b, where the achieved gain is shown and the curve resulted through the simulation of the reservation mechanism is not distinguishable from the one produced by the LFMIS. From this figure, we can also notice that there is a small, but visible difference between the analytical results of the reservation mechanism and those obtained through simulation, as the clique assumption used in the analytical framework leads to an under-estimation of the number of uplink users. Nevertheless, both analytical and simulation results show that RB is able to discover, under high vehicular density, a transmitter set containing around 20% less nodes than the solution obtained using DB-C, and close to an MDS obtained by centralized algorithms.

In a vehicular network, there are several possibilities for transmitting the reservation information. The one we consider preferable consist in transmitting short reservation messages on the service channel used by non-safety applications, where reservation information could also be piggy-backed in other non-safety messages. In this case, we consider that reservation periods containing several hundred slots are feasible even when transmissions on the cellular uplink happen with a granularity in the order of seconds, therefore close to optimal solutions are achievable under realistic assumptions.

Two other elements can impact in practice the performance of the RB mechanism: radio propagation problems and node mobility. Indeed, a reservation message might not reach all the neighbors, meaning that some covered nodes might not cancel their back-off, and result in an unnecessary transmission on the cellular uplink. On the other hand, it is also possible for a node to receive a reservation message from a node situated at a distance higher than the threshold R used to define the one-hop neighborhood in Sec. II-B. While vehicles can simply filter these messages coming from outside the reservation range, handling propagation problems would require additional transmissions, hence higher reservation periods. The model presented in this section can be easily extended to take into account radio propagation issues by including the link failure probability in Eq. 8.

Mobility can also affect the RB mechanism, as a vehicle can move in an area already covered by a vehicle after the reservation message has been sent. The vehicle in question would find itself between two dominator nodes but without being covered by any of them. This situation would once again result in an unnecessary transmission on the cellular network. However, given that vehicles move of a few meters at most during a reservation interval, we observed mobility to have marginal impact on the offload performance.

V. CONCLUSIONS

Floating Car Data are envisioned to enable a number of applications for intelligent transportation and urban sensing. However, the proliferation of FCD-based services might induce high uplink load on the cellular networks. In this paper, we propose a first study of FCD offload through vehicular communication. **In the optimal case, this approach is able to offload up to 95% of the FCD from the cellular uplink at peak hours**. Moreover, we design and analyze three very simple distributed heuristics, and **we show that simple practical schemes can achieve near-optimal performance with no actual calibration required.** Overall, our results indicate that V2V-based local gathering and fusion of FCD could significantly reduce the demand for uplink bandwidth at the access network.

REFERENCES

[1] R.N. Charette, *"This car runs on code"*, IEEE Spectrum, February 2009.
[2] U. Paul, A.P. Subramanian, M.M. Buddhikot, S.R. Das, *"Understanding traffic dynamics in cellular data networks"*, IEEE INFOCOM, Shanghai, China, April 2011.
[3] Cisco, *"Global Mobile Data Traffic Forecast Update, 20112016"*, White Paper, February 2012.
[4] C. de Fabritiis, R. Ragona, G. Valenti, *"Traffic Estimation And Prediction Based On Real Time Floating Car Data"*, IEEE ITSC, Beijing, PRC, October 2008.
[5] S. Eichler, *"Anonymous and Authenticated Data Provisioning for Floating Car Data Systems"*, IEEE ICCS, Singapore, October 2006.
[6] J. Whitbeck, Y. Lopez, J. Leguay, V. Conan, M. Dias de Amorim, *"Relieving the wireless infrastructure: when opportunistic networks meet guaranteed delays"*, IEEE WoWMoM, Lucca, Italy, June 2011.
[7] B. Han, P. Hui, V.S.A. Kumar, M.V. Marathe, G. Pei, A. Srinivasan, *"Cellular traffic offloading through opportunistic communications: a case study"*, ACM CHANTS, Chicago, IL, USA, September 2010.
[8] F. Malandrino, C. Casetti, C.-F. Chiasserini, M. Fiore, *"Offloading Cellular Networks through ITS Content Download"*, IEEE SECON, Seoul, Korea, June 2012.
[9] S. Dimatteo, P. Hui, B. Han, V.O.K. Li, *"Cellular Traffic Offloading through WiFi Networks"*, IEEE MASS, Valencia, Spain, October 2011.
[10] G. Hertkorn, P. Wagner, *"The application of microscopic activity based travel demand modelling in large scale simulations"*, World Conference on Transport Research, Istanbul, Turkey, July 2004.
[11] G. Rindsfüser, J. Ansorge, H. Mühlhans, *"Aktivitätenvorhaben"*, in K.J. Beckmann (editor), SimVV Mobilität verstehen und lenken – zu einer integrierten quantitativen Gesamtsicht und Mikrosimulation von Verkehr, Final report, Ministry of School, Science and Research of Nordrhein-Westfalen, Düsseldorf, Germany, 2002.
[12] C. Gawron, *"An Iterative Algorithm to Determine the Dynamic User Equilibrium in a Traffic Simulation Model"*, International Journal of Modern Physics C, 9(3):393–407, 1998.
[13] S. Uppoor, M. Fiore, "Large-scale Urban Vehicular Mobility for Networking Research", *IEEE VNC*, Amsterdam, Holland, November 2011.
[14] F. Bai, D.D. Stancil, H. Krishnan, "Toward Understanding Characteristics of Dedicated Short Range Communications (DSRC) From a Perspective of Vehicular Network Engineers", *ACM MobiCom*, Chicago, IL, USA, September 2010.
[15] S Boccaletti, V. Latora, Y. Moreno, M. Chavez, D. Hwang, "Complex Networks: Structure and Dynamics", *Elsevier Physics Reports*, 424(4-5):175–308, January 2006.
[16] F. Kuhn, T. Moscibroda, R. Wattenhofer, "The Price of Being Near-Sighted", *ACM-SIAM SODA*, Miami, FL, USA, January 2006.
[17] M. Luby, "A Simple Parallel Algorithm for the Maximal Independent Set Problem", *ACM STOC*, Providence, RI, USA, May 1985.
[18] L. Lovasz, *"On the Ratio of Optimal Integral and Fractional Covers"*, Discrete Mathematics, 13(4):383–390, December 1975.
[19] M. Marathe, H. Breu, H. Hunt III, S. Ravi, D. Rosenkrantz, *"Simple Heuristics for Unit Disk Graphs"*, Networks, 25(2):59–68, March 1995.
[20] H. Hunt III, M. Marathe, V. Radhakrishan, S. Ravi, D. Rosenkrantz, R. Stearns, *"NC-Approximation Schemes for NP- and PSPACE-Hard Problems for Geometric Graphs"*, Journal of Algorithms, 26(2):238–274, February 1998.
[21] S.P. Boyd, L. Vandenberghe, *"Convex Optimization"*, Cambridge University Press, 2004.
[22] R.P. Brent, *"Algorithms for Minimization Without Derivatives"*, Dover Publications, 2002.