HAL
open science

Performance Boundaries of Massive Floating Car Data Offloading

Silvia Ancona, Razvan Stanica, Marco Fiore

To cite this version:

Silvia Ancona, Razvan Stanica, Marco Fiore. Performance Boundaries of Massive Floating Car Data Offloading. WONS - 11th IEEE/IFIP Annual Conference on Wireless On-demand Network Systems and Services, Apr 2014, Obergurgl, Austria. pp.89 - 96, 10.1109/WONS.2014.6814727 . hal-01005015

HAL Id: hal-01005015

https://inria.hal.science/hal-01005015v1

Submitted on 11 Jun 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Performance Boundaries of Massive Floating Car Data Offloading

Silvia Ancona[1,3] and Razvan Stanica[1] and Marco Fiore[1,2]

[1]*INSA Lyon, CITI-INRIA UrbaNet, Villeurbanne, France*
[2]*CNR-IEIIT, Torino, Italy*
[3]*Politecnico di Bari, Bari, Italy*

Floating Car Data (FCD) consist of information generated by moving vehicles and uploaded to Internet-based control centers for processing and analysis. As upcoming mobile services based on or built for networked vehicles largely rely on uplink transfers of small-sized but high-frequency messages, FCD traffic is expected to become increasingly common in the next few years. Presently, FCD are managed through a traditional cellular network paradigm : however, the scalability of such a model is unclear in the face of massive FCD upload, involving large fractions of the vehicles over short time intervals. In this paper, we explore the use of vehicle-to-vehicle (V2V) communication to partially relieve the cellular infrastructure from FCD traffic. Specifically, we study the performance boundaries of such a FCD offloading approach in presence of best- and worst-case data aggregation possibilities at vehicles. We show the gain that can be obtained by offloading FCD via vehicular communication, and propose a simple distributed heuristic that has nearly optimal performance under any FCD aggregation model.

Keywords: vehicular networks, floating car data, cellular uplink

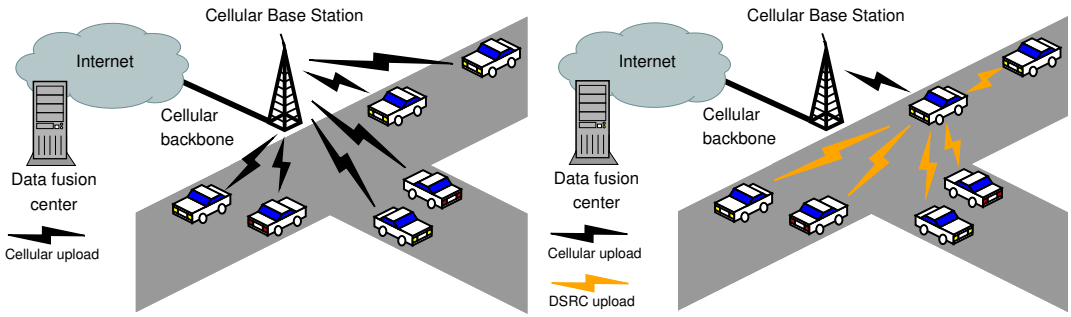
1 Introduction

The term Floating Car Data (FCD) defines generic information collected by cars – privately owned vehicles as well as public transportation means – while traveling to their destination, and uploaded to Internet-based control centers for processing and analysis. Since FCD-based technologies have started to become commercially appealing only during the last few years, their diffusion is still limited. Yet, they already significantly improved mobility-critical services, and FCD is today employed for different practical applications.

One example is that of *distant monitoring of vehicles*. Vehicles are today equipped with a wide range of interconnected microprocessor-based Electronic Control Units (ECUs), whose number varies between 30 for low-end cars and 100 for premium-class automobiles [1]. Many of these ECUs are in charge of monitoring and possibly reacting to information collected by on-board sensors, that constantly gather data on the vehicle mechanics, the driver and passenger space and the external environment. The use of FCD generated by the ECUs to locally oversee in-vehicle operations has been a standard practice since the 90's, and today just a few car functionalities are not under control of ECUs. Recently, car manufacturers have started to propose remote monitoring of the car sensors, by exploiting the cellular connectivity to retrieve the ECU-generated FCD and provide seamless distant support to the driver and the passengers. Systems such as BMW Assist, Ford SYNC, General Motor OnStar, Toyota Safety Connect and Mercedes-Benz mbrace, just to cite a few representative examples, include safety, diagnostic and anti-theft applications.

Another example of FCD-based service is *real-time traffic control*. High-end car navigation systems embed cellular interfaces that are used to upload FCD to traffic control centers, in the form of anonymized vehicle speed and position information, obtained through the GPS receiver. This data is processed to determine the conditions of road traffic in real-time and suggest more efficient routes to the drivers. Examples of systems leveraging FCD are, e.g., TomTom HD Traffic and Meihui TrafficCast.

These applications only represent the tip of the iceberg, as many future automotive solutions are foreseen to heavily rely on FCD. The most significative example is probably that of *urban sensing*, regarded as a



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

FIGURE 1: FCD upload scenarios : traditional and offloaded.

fundamental component of forthcoming smart cities. Urban sensing is envisioned to enable a fine-grained yet pervasive and real-time monitoring of critical factors in large urban areas, such as mobility patterns, pollution levels, resource usages. To that end, cars are expected to continuously collect FCD on the metropolitan environment they travel in, and to upload it to Internet-based control centers for data fusion and knowledge discovery.

We argue that the consolidation and diffusion of FCD-based technologies may easily reach very high penetration rates in the next few years. Also, ever-growing FCD upload frequencies and volumes are to be expected, as more fine-grained measures are required and as the number of features to be monitored (both inside and outside cars) increases. Overall, we could soon witness the emergence of *massive FCD* traffic.

From a communication infrastructure viewpoint, massive FCD would call the present network model into question. FCD are currently gathered by moving vehicles and uploaded to postprocessing centers through the cellular infrastructure. To that end, the service providers sign agreements with the telecom operators, so that FCD transfers occur through the cellular infrastructure in a way that is completely transparent to the user, i.e., the vehicle's driver or passengers.

Such a communication paradigm is sustainable at the negligible penetration rates and low update frequencies of today's FCD-based technologies. However, the appearance of FCD will induce a dramatic increase in the cellular network uplink load. Today's mobile access infrastructure is already at its capacity limit [2] and forecasts anticipate that mobile traffic to outgrow the access network capacity 5 to 1 by 2018 [3]. Massive FCD risk to be an additional heavy burden, when picturing scenarios where the hundreds of thousands of vehicles that normally circulate in major metropolitan areas during rush hours all try to obtain uplink access to the cellular network every few seconds, to transmit a several hundreds of KB of information at each access. Even more so, when thinking that traditional offloading via Wi-Fi or femtocells – that relieved the cellular network from 45% of the total mobile traffic in 2013 [3] – assumes low user mobility and is thus not viable in the FCD case.

In order to prevent FCD from seizing the cellular capacity, new offloading techniques are to be devised, which scale well to citywide scenarios. In this paper, we analyse a FCD offloading paradigm that leverages direct vehicle-to-vehicle (V2V) communication based on Dedicated Short Range Communication (DSRC). DRSC has been the subject of a quite active standardization activity in recent years, leading to a number of protocol stack proposals, including IEEE 802.11p, IEEE 1609.x, ETSI ITS G5 and ISO CALM. These standards enable communication among vehicles traveling at a few hundreds of meters distance. The offloading approach is exemplified in Fig. 1. In the traditional model, in Fig. 1a, each vehicle uploads the FCD it collects independently. Conversely, the V2V-based FCD offloading model, in Fig. 1b, commends that DSRC communication is employed to convey FCD to a subset of selected vehicles, which gather and aggregate the data their receive before uploading it via the cellular network.

The system could result in significant offload of the cellular infrastructure, in terms of channel setup signalization, control header overhead and, depending on the local aggregation levels, sheer uplink traffic volume. In this work, we focus on the latter aspect, i.e., the gain that V2V-based FCD offload can achieve under different local aggregation models. By local aggregation, we mean how much vehicles can compress

the data their receive from their DSRC communication neighbors, before uploading them via the cellular infrastructure. Clearly, the more the FCD can be aggregated locally at the uploader vehicles, the lower the traffic load to be sustained by the cellular access network. In particular, we investigate the performance boundaries of V2V-based offload, i.e., evaluate its gain when the possibilities for local aggregation are maximum or minimum, respectively.

To that end, after a brief review of the literature in Sec. 2, we present in Sec. 3 the system model we consider, describing the best- and worst-case local aggregation functions, and the scenario used in our study. Sec. 4 and Sec. 5 present centralized and distributed solutions, respectively, and discuss the corresponding performance. Finally, Sec. 6 draws conclusions and outlines future research directions.

2 Related work

The V2V-based FCD offload system we study lies at the interface of FCD management and mobile offloading.

Concerning the first topic, FCD management is a mainstream research subject mainly in traffic planning theory [4–6] and in knowledge discovery [7]. From a networking perspective, it has been originally tackled in terms of user privacy [8, 9]. More recent works have started considering mobile access as well, studying how to accomodate FCD through traditional cellular upload [10, 11] These works have focused on exploitation of the Random Access Channel (RACH) of UMTS access networks, which is a shared direct-access channel normally employed for signalization. By sending small, frequent data directly on the RACH, vehicles avoid to ask dedicated resources just to upload a few bytes. Results show that the solution is viable, but also that important issues exist : FCD uploads via RACH are quite limited in number (order of a few tens per second within each cell), suffer from the lack of handover over the RACH, and may severely reduce the QoS of traditional mobile voice and data services by congesting the RACH. LTE is instead considered as the access network technology in [12], where the impact of FCD upload on energy consumption and traditional mobile traffic is assessed. However, the analysis does not consider massive FCD, assuming that 5% of vehicles uploading FCD at a frequency adjusted so as limit the blocking probability of traditional traffic. Finally, a comparison of UMTS and LTE technologies for the dissemination of cooperative awareness messages (CAMs) – a process which includes upload of FCD on RACH – was performed in [13]. Results showed that LTE outperforms UMTS in terms of FCD upload capabilities, but also outlines that DSRC remains the fittest technology for CAM dissemination. As a matter of fact, LTE also suffers from higher delays (up to 50 ms) and limited FCD upload capacity (up to 100 messages per second per 5-MHz channel) which is also negatively affected by the presence of traditional mobile traffic.

The previous literature on cellular network offloading has instead mainly focused on downlink data transfers, targeting the dissemination of content to smartphone users [14] or car passengers [15] and the download of large contents by vehicular users [16]. 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. 1, typically require the upload to occur in quasi real-time.

A few recent works have started considering the two topics at once, proposing solutions to offload FCD through DSRC communication. Specifically, they considered using transfer opportunities among vehicles [18] as well as between vehicles and dedicate roadside units (RSUs) [17]. In both cases, load reductions at the cellular access infrastructure in the order of 90% have been reported. Despite their seminal contribution, such works have disregarded the local aggregation aspect [18], or considered low FCD upload frequencies, and scenarios that are limited geographically and temporally [17]. In our work, we focus on massive FCD, and evaluate the impact of local aggregation in a large-scale simulation environment.

3 System Model

We model the vehicular network as a temporal graph $G(\mathbb{V}(t), \mathbb{E}(t))$, where $\mathbb{V}(t) = \{v_i\}$ is the set of nodes at time t, and $\mathbb{E}(t) = \{e_{ij}(t) \,|\, v_i, v_j \in \mathbb{V}, i \neq j\}$ is the set of edges at the same time instant. Each node v_i maps to a vehicle traveling in the considered scenario, and an edge $e_{ij}(t)$ exists if vehicle i and vehicle j can act as local FCD aggregators for one another at time t. Specifically, in the following we consider a simple unit disk

graph model to represent aggregation capability, implying that information regarding vehicle i is known by all other vehicles in a range R. We point out that such a model does not make any assumption regarding the radio propagation model : it simply means that application-level information can be gathered from a certain area during a collection period determined by the service requirements, using direct or multi-hop DSRC-based V2V communication.

Indeed, different classes of applications can be defined, based on the frequency and accuracy required to the FCD upload process. For example, a real-time traffic information system, where the traveling time is estimated at the scale of a city – if not at that of a region or of a whole country – uses information regarding the speed of the vehicles on different road segments. However, individual measures are noisy and redundant, and average measures are needed. This means that a local aggregation is possible within the vehicular network, since a vehicle can gather speed information from several neighbors and upload only an average value. Similar considerations can be made for other applications allowing for FCD local aggregation, such as environmental monitoring, where pollution or temperature sensors embedded in vehicles are used to build a dynamic map of the phenomena over a certain geographical area.

While interesting from a networking and privacy preservation point of view, in-network aggregation is not always possible. For example, taxi or bus companies need to monitor the vehicles in their fleet individually, for logistic and maintenance purposes. Insurance companies also propose services for which individual vehicle information is required. As disscussed in detail later on, in this cases, reducing the number of uploaders can still bring benefits, which are however reduced by the lack of local aggregation.

In this work, we concentrate on the advantages that DSRC-based FCD offload can bring in terms of the amount of data uploaded on the cellular uplink. That perspective is especially useful to private vehicular users or fleet operators since FCD traffic volumes can be directly translated to a financial cost. We consider instead that messages exchanged directly between vehicles, using DSRC technologies, do not incur in any cost, since they operate on non-commercial frequency bands. However, we also pay close attention to the DSRC overhead introduced by the different offload solutions, so as not to neglect the limited capacity of the V2V broadcast channel.

3.1 FCD upload and local aggregation models

At time t, representing the end of an FCD collection period, a set of vehicles $\mathbb{U}(t) \subseteq \mathbb{V}(t)$ use the cellular uplink to transmit data. The FCD upload cost for that collection period can thus be expressed as :

$$C(t) = \sum_{v_i \in \mathbb{U}(t)} c(v_i)$$

where $c(v_i)$ represents the amount of data transmitted on the cellular network by vehicle v_i.

In the baseline case, vehicles do not collaborate using V2V communication, each uploading its own data via cellular access. Therefore, $\mathbb{U}(t) = \mathbb{V}(t)$, and each vehicle needs to upload its own application-layer payload of p bytes that, cumulated with network- and transport-layer headers† totalling h bytes, results in a message of a total size $m = h + p$. Since we consider that size to be constant over vehicles and collection peridos, $c(v_i) = m$, $\forall v_i \in \mathbb{U}(t), \forall t$, and the FCD upload cost in the baseline case is :

$$C_{base}(t) = \sum_{v_i \in \mathbb{V}(t)} c(v_i) = m \cdot |\mathbb{V}(t)|.$$

If FCD can instead be offloaded through direct vehicular communications, the number of uploaders is no longer equal to the number of nodes, as some vehicles act as uploaders also for FCD collected by a part of their neighbors. In that case, $\mathbb{U}(t)$ is normally a proper (and significantly smaller) subset of $\mathbb{V}(t)$. Clearly, the smaller the number of uploaders, the higher the offload gain brought by DSRC-based communication. Therefore, how $\mathbb{U}(t)$ is determined represents a key aspect of the problem, and it will be the focus of Sec. 4 and Sec. 5. Also, in presence of DSRC-based FCD offload, the individual cost $c(v_i)$ depends on the aggregation function that can be used by relay vehicles. Many such functions can be imagined, but in the

†. A recent experimental study shows that around 75% of cellular operators account for traffic at an IP level [19], meaning that not only the payload, but also the headers need to be considered in the cost computation.

following we focus our analysis on two boundary cases, that we denote as *best-case* and *worst-case* local aggregation models, respectively.

Best-case aggregation. In the best-case aggregation scenario, individual information for each vehicle is not required. A node can therefore gather FCD from all its $N_i - 1$ aggregation neighbors ‡ and compute a single statistical measure (e.g., an average) of the values carried by all the received payloads. The gain in this case is significant : instead of uploading N_i separate messages of size m as in the baseline case, each uploader only sends one message of the same size m that summarizes the FCD of its entire aggregation neighborhood. That maps to the maximum level of in-network aggregation. The resulting cost is :

$$C_{best}(t) = \sum_{v_i \in \mathbb{U}(t)} c(v_i) = m \cdot |\mathbb{U}(t)|.$$

Worst-case aggregation. In the worst-case scenario, we consider that individual information is required for every vehicle. This means that a node gathering data from $N_i - 1$ neighbors needs to upload N_i distinct payloads, i.e., that no local aggregation is possible. However, there is still some space for improvement over the baseline case above : FCD payloads from different vehicles can be fit in the same packet, resulting in lower overhead in terms of network- and transport-layer headers. As FCD-based services usually require minimal measures, headers can account for a significant part of the upload cost. Payload stuffing allows an uploader vehicle v_i to only transmits R_i packets instead of N_i, with :

$$R_i = \left\lceil \frac{N_i \cdot p}{m_{max} - h} \right\rceil ,$$

where m_{max} is the maximum size authorized for a network-layer packet. The resulting cost for the worst-case aggregation at a generic time t is then :

$$C_{worst}(t) = \sum_{v_i \in \mathbb{U}(t)} c(v_i) = \sum_{v_i \in \mathbb{U}(t)} (R_i \cdot h + N_i \cdot p).$$

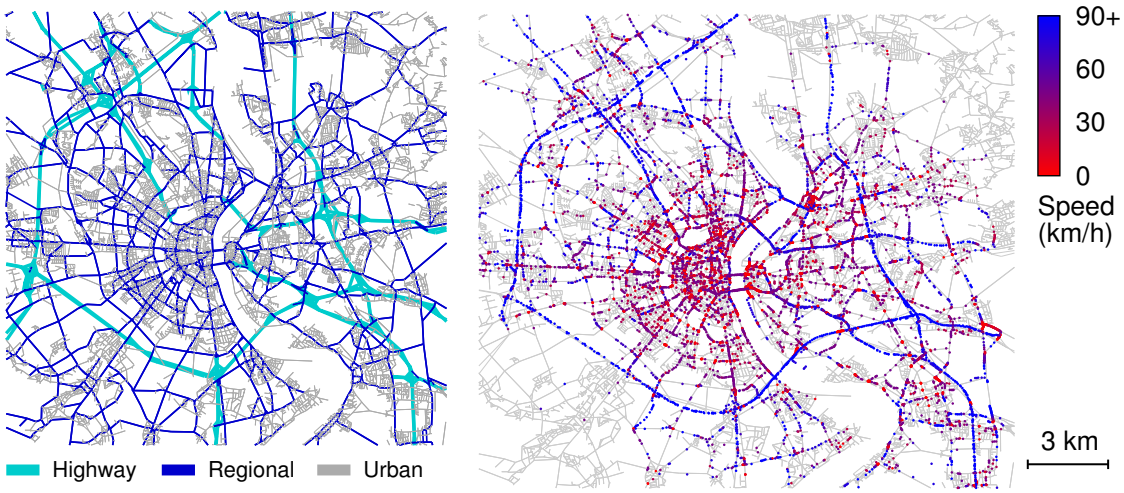
It is noteworthy that having a reduced number of users on the cellular uplink also brings benefits in terms of network access delay and availability. However, we do not consider the impact of these parameters on the cost, as they do not directly influence the economic cost of the FCD upload from a end user's standpoint.

3.2 Reference scenario

The reference scenario we will consider in this study is that of the metropolitan region of Cologne, Germany. The scenario comprises a 400 km^2 surface that includes the city of Cologne as well as the surrounding suburban and rural areas. Approximately 4.500 km of roads are present in the region, including major highways, urban arteries and minor streets. The layout of the road network in the reference scenario is outlined in Fig. 2a.

Road traffic in the region is provided by a state-of-art vehicular mobility dataset that describes 24 hours of individual car movements during a typical working day, and includes more than 700.000 trips within and across the Cologne area. The dataset was generated by coupling different tools for the road topology information, microscopic mobility modeling and macroscopic traffic flow description. 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. On the one hand, the Travel and Activity PAtterns Simulation (TAPAS) methodology [20] is applied on real-world data collected in the Köln region [21] 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 [22] is run to determine a traffic assignment (i.e., the routes taken by each driver) that allows a so-called dynamic user equilibrium.

‡. We cosider that N_i is the size of the local aggregation neighborhood of a node, i.e., the number of nodes for which v_i can act as aggregator and single uploader. Therefore v_i itself is counted in N_i, and the number of actual aggregation neighbors is $N_i - 1$.



(a) Road network layout

(b) Road traffic snapshot

FIGURE 2: Road network layout and road traffic at 7 am in the reference scenario of Cologne, Germany.

A snapshot of road traffic in the Cologne, as described by the dataset at 7 am, is portrayed in Fig. 2b. There, each dot represents one car, its color corresponding to its speed : we can remark the high-speed highways around the conurbation, and more congested traffic in the city center. For further details on the dataset, we refer the reader to [23].

4 Centralized DSRC-based FCD offload

As anticipated in Sec. 3.1, a key operation in the FCD offload process is that of identifying the subset of vehicles $\mathbb{U}(t)$ in charge of performing the FCD local aggregation and upload at a given time instant t. The goal to select $\mathbb{U}(t)$ so as to minimize the upload cost (i.e., maximize the offload gain) while guaranteeing that FCD is collected from all vehicles participating in the system (i.e., without affecting the quality of the FCD collection).

In this section, we focus on centralized solutions to such a problem. These take an oracle approach, assuming that the $G(\mathbb{V}(t),\mathbb{E}(t))$ structure is known.

4.1 Algorithms

In [18], the optimal uploader set selection problem is shown to be equivalent to the **minimum dominating set (MDS)** problem in graph theory, known to be NP-complete. However, this equivalence is true if we use as criterion the number of vehicles using the cellular uplink, but does not hold when the optimization is focused on the actual amount of data transferred to the cellular network.

To better understand the difference, we recall that a *dominating set* over a set $\mathbb{V}$ of nodes is defined as $\mathbb{S} = \{v_i \mid \exists\, e_{ij}\ \forall v_j \in \mathbb{V} \setminus \mathbb{S}\}$. In our case, this means that each vehicle is either in the dominating set, or has an aggregation neighbor in this set. The dominating set is said to be *minimal* if removing any node from the set breaks the dominance property, and the MDS is the minimal dominating set with the smallest size.

When focusing on the sheer number of uploaders, the MDS of the graph $G(\mathbb{V}(t),\mathbb{E}(t))$ is by definition the optimal solution. However, in an MDS, a node can be dominated by one or multiple (dominator) neighbors. In the best-case aggregation scenario, this difference is not important, since each dominator vehicle only uploads average values computed over its entire neighborhood. However, in the worst-case scenario, individual information is required from each vehicle : this implies that each dominator must upload all the FCD payloads it collects, thus leading to redundant transmissions. Indeed, the FCD generated by a vehicle v_i with K dominating neighbors, $v_1 \dots v_K$, will be transmitted K times, once by each dominator.

To solve this problem, we propose three centralized heuristics to compute the set of vehicular relays considering the costs we defined in the previous section. In the first approach, denoted as **minimum dominating**

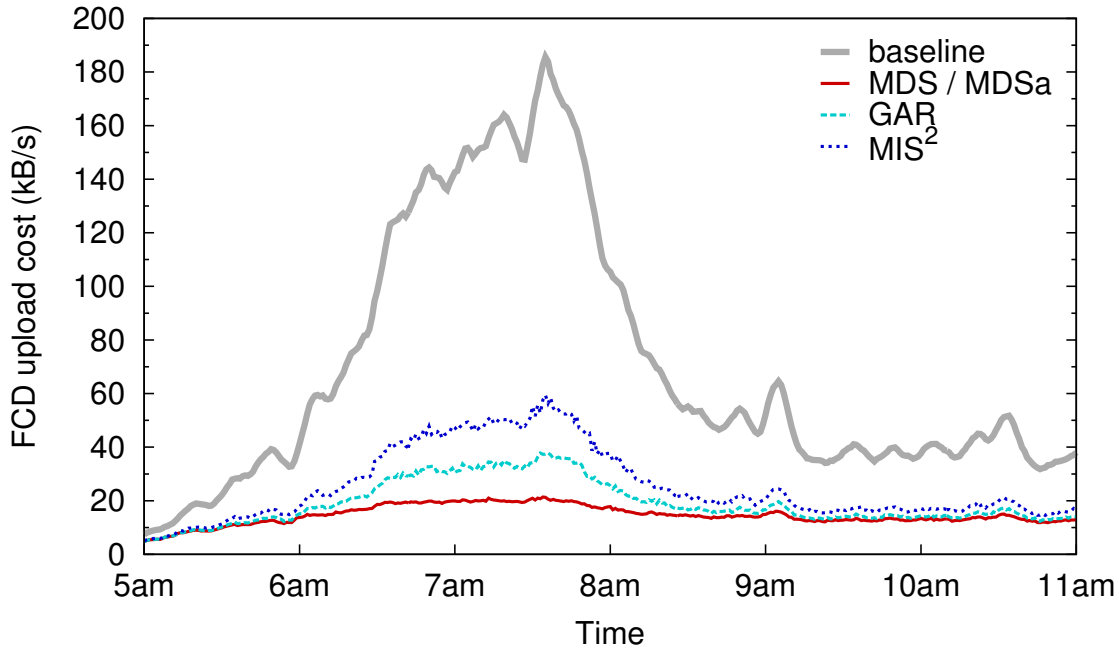


FIGURE 3: Cost of uploading FCD in the best-case local aggregation scenario using centralized approaches.

set with assignment (MDSa), we compute an approximate solution for the MDS problem (remember the problem is NP-complete) using the greedy algorithm in [18]. This algorithm starts with an empty set $\mathbb{S}$ and iteratively adds to it the node with the highest number of neighbors outside $\mathbb{S}$, which results in a dominating set in a finite number of iterations. However, we add a second step to this algorithm, by assigning all the nodes that remain outside $\mathbb{S}$ to only one node from the dominating set. This practically means that a vehicle dominated by two neighbors only transmits the required information to one of them.

In the second proposed solution, **greedy assignment with removal (GAR)**, we modify the MDS greedy algorithm described above, by picking at each iteration only nodes that have no neighbors in $\mathbb{S}$. This algorithm does not result in a dominating set, as some nodes remain not dominated after the last iteration. The vehicles in this situation will transmit their own FCD (but not their neighbors') on the cellular uplink.

Finally, the third approach consists in computing a **maximal independent set on the square of the original graph (MIS^2)**. An independent set contains only nodes that are not connected by an edge, and it is called *maximal* if no other node from the graph can be added to the set without breaking the independence property. Just as the complementary MDS problem, computing the maximum independent set (MIS) is NP-complete and approximation algorithms need to be used. An important property is that nodes in an MIS are separated by a minimum distance of two (aggregation) hops. In our system, if we want any node to have at most one neighbor in the uploading set, the nodes in the independent set need to be situated at minimum three hops from one another. In order to achieve this separation, we build a maximal independent set on a graph where a node is connected by an edge to all its one-hop and two-hop neighbors, i.e. the square of the graph. However, just like in the case of GAR, this algorithm does not guarantee covering all the nodes in the original graph, and the remaining vehicles need to upload their own FCD.

4.2 Results

We tested the performance of the centalized algorithms on a 6-hour timespan during the morning of the Cologne reference scenario presented in Sec. 3.2. We remark that such a time period is fully representative of different vehicular traffic intensities, as it includes sparse (5 am - 6 am), peak (6 am - 8 am), and moderate (8 am - 11 am) traffic conditions. Unless mentioned otherwise, we consider TCP/IP headers with a size $h = 20$ bytes (we make the assumption of a Robust Header Compression mechanism such as the one used in LTE cellular networks [24]), a payload $p = 100$ bytes, and the maximum size of an IP packet $m_{max} = 1440$ bytes.

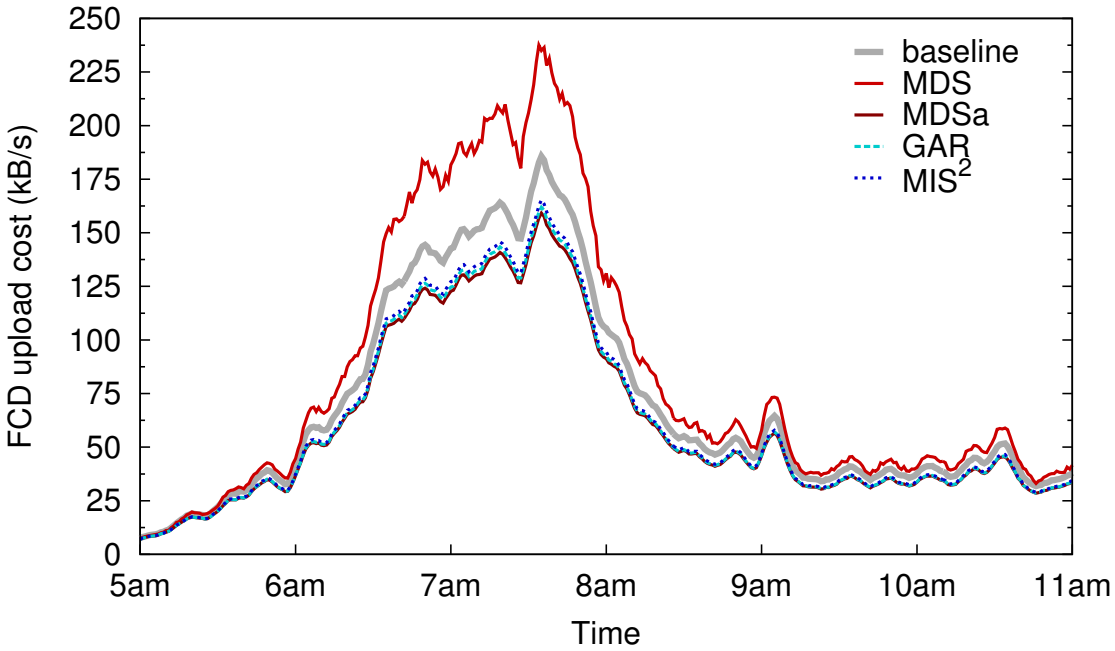


FIGURE 4: Cost of uploading FCD in the worst-case local aggregation scenario using centralized approaches.

We set the collection period to 10 seconds, resulting in an upload frequency 0.1 Hz, and assumes that the range within which vehicles can locally aggregate FCD is $R = 100$ m.

In Fig. 3, we show the cost of uploading FCD in the case of applications that allow a maximum degree of local aggregation. As in this case the information from all the neighbors is aggregated in a single message of size m, the offloading performance is directly given by the number of vehicular relays, therefore the MDS and MDSa algorithms, that pick the same uploaders set, result in identical performance. Compared with a non-collaborative approach, an MDS-based offloading significantly reduces the upload cost from a peak of around 180 kB/s to less than 20 kB/s. The two other approaches, GAR and MIS^2 are less efficient than MDSa, but still manage to reduce the cost by more than 50% even in medium density traffic.

However, when we consider the worst case aggregation scenario, the results presented in Fig. 4 show that the cost of the original MDS algorithm is even high than that of the non-offloaded baseline approach, especially at peak hours. This is due to the fact that redundant information is uploaded in the case of vehicles with multiple neighbors in the MDS. Nevertheless, the three heuristics proposed above manage to achieve a cost reduction, by decreasing the number of transmitted TCP/IP headers. In high density traffic, between 6am and 8am, a gain of around 20 kB/s can still be obtained by offloading FCD when no local aggregation is possible.

In such worst-case aggregation scenario, the ratio between the header size and the payload size becomes an important parameter. In Fig. 5, we keep a constant header size $h = 20$ bytes, and we vary the payload size between 20 and 500 bytes. For payloads comparable with the size of the TCP/IP headers, even the basic greedy MDS algorithm manages a 20% cost reduction compared with a scenario where all the vehicles use the cellular uplink. However, the MDS cost quickly increases with the payload size, and unnecessary transmissions make this solution more costly than a non-collaborative approach. The other three algorithms always outperform the reference solution, with MDSa giving the best results, and more than 40% cost reduction for a payload of 20 bytes.

To summarize, these results show that **offloading FCD can bring a significant cost decrease** from the users' point of view, **especially when the application allows a high degree of data aggregation**. Moreover, even when data aggregation is not possible, **an important gain can be achieved if the header-to-payload ratio is high**.

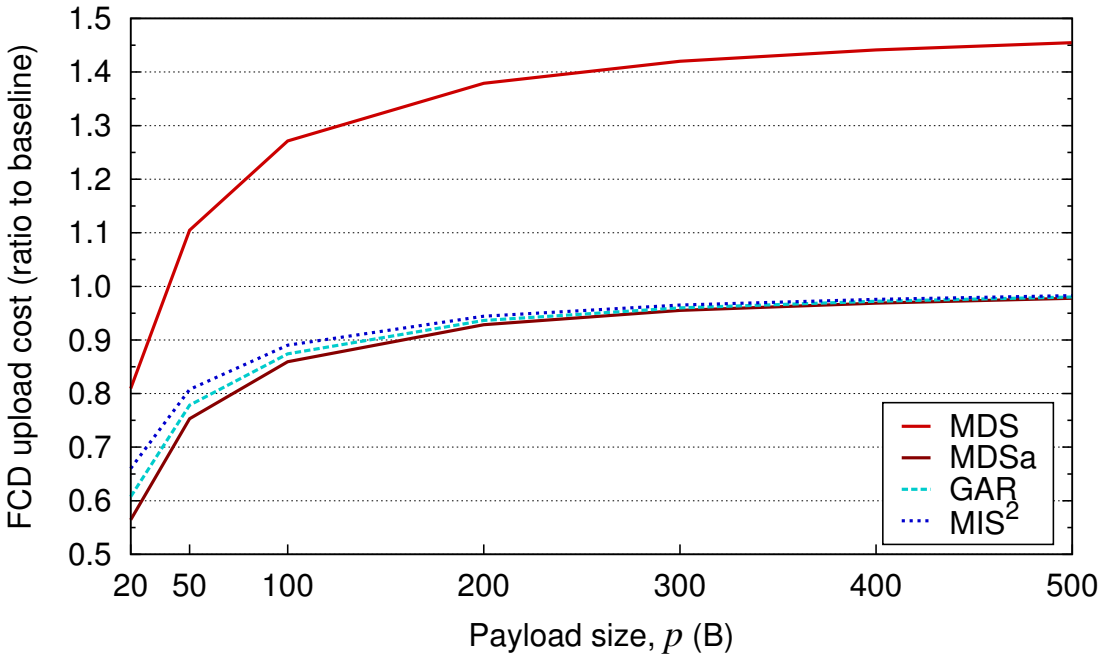


FIGURE 5: Relative cost of uploading FCD (with respect to baseline scenario with no offload) for the worst case aggregation function in the case of centralized algorithms. Results are averaged between 5am and 11am.

5 Distributed DSRC-based FCD offload

The centralized algorithms presented in the previous section give an idea on the cost gains that can be brought by offloading FCD through vehicle-to-vehicle communication. In this section, we check whether practical, distributed solutions can also reduce the FCD upload costs.

5.1 Algorithms

Three distributed algorithms for DSRC-based FCD offload have been proposed in [18], as follows.

In the **degree-based (DB)** algorithm, a vehicle acts as a relay with a probability that linearly decreases with the number of neighbors. This method does not add any overhead in the V2V network, but it also does not give any guarantee on gathering information from all the vehicles in the network.

The **degree-based with confirmation (DBC)** algorithm adds an extra step to DB, by requiring every relay vehicle to also confirm its state by transmitting a message on the vehicular network. At the end of the collection period, the nodes that were not selected as relays and did not hear any confirmation message from one of the neighbors upload their own FCD on the cellular network.

Finally, the **reservation-based (RB)** algorithm introduces a reservation phase at the beginning of every collection period. In this phase, all the vehicles randomly choose a slot in which they declare themselves as relays if they have not previously received such a message from a neighbor.

The goal of all these algorithms is to approximate an MDS that, as discussed in Sec. 4, is an optimal solution only when massive data aggregation is allowed. To take into consideration applications that do not allow FCD aggregation, we propose an additional distributed algorithm, named **reservation-based with confirmation (RBC)**. This new solution builds on the reservation phase idea used in RB, requiring every vehicle to randomly select a slot from this time frame. However, unlike the original approach, RBC provides feedback to the vehicles that win the contention and act as relays. Indeed, in RB relay nodes upload FCD from all their neighbors, meaning that information from vehicles with two neighboring relays will be transmitted on the cellular network twice.

To counter that effect, RBC uses two types of messages : *relay selection (RS)* and *relay confirmation*

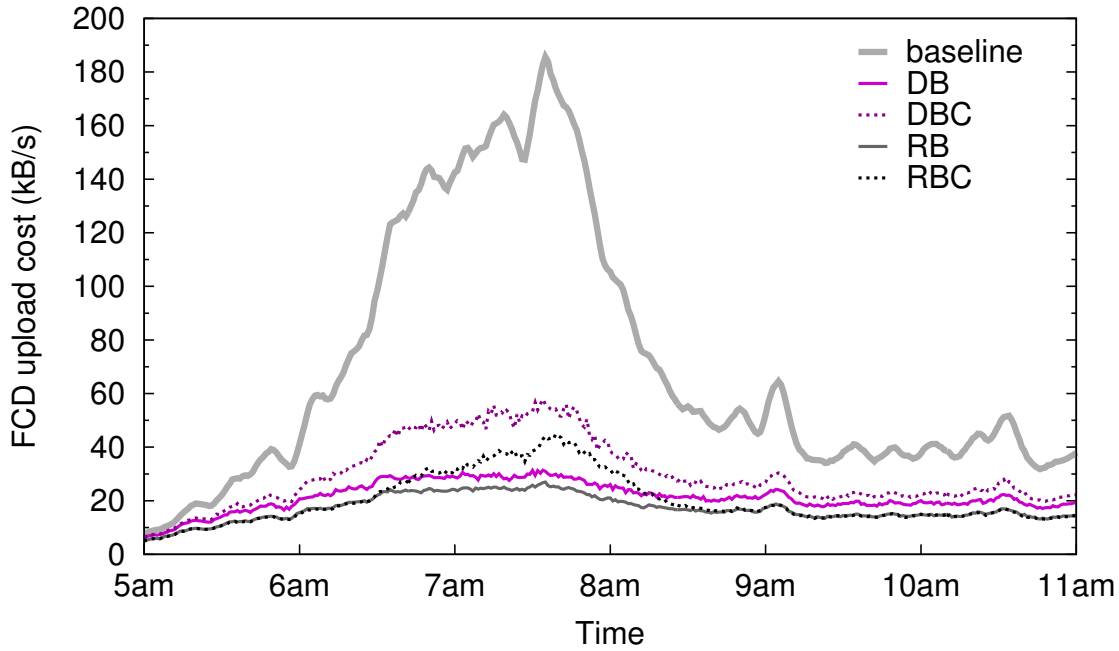


FIGURE 6: Cost of uploading FCD in the best-case local aggregation scenario using distributed approaches.

(RC). In the randomly selected slot, a vehicles transmits either an RS or an RC, as follows : (i) if the vehicle did not receive an RS message from one of its neighbors in the previous slots, it declares itself as a relay and transmits RS ; (ii) if the vehicle previously received an RS message from one of its neighbors, it transmits instead an RC message, confirming the neighbor in question as a relay. When a vehicle is covered by multiple relays, it randomly picks one of them using an RC message. At the end of the collection phase, a vehicle acting as a relay only uploads FCD coming from neighbors that confirmed it as a relay.

In the ideal case, RBC assigns all the vehicles to one relay and assures a minimum distance of two aggregation hops between any two relays. However, in a real scenario, the number of slots in the reservation phase is limited, meaning that RS and RC messages can collide, resulting in sub-optimal performance. Some vehicles may not receive RS messages transmitted by their neighbors and declare themselves as uploaders without this being needed. Also, some uploaders might lose some of their neighbors' RC messages due to collisions, which means the neighbors in question will need to upload their own information despite having a close-by uploader. To test the performance of RBC, we take into consideration these collisions in the case where reservation slot access is regulated by a simple slotted-Aloha mechanism.

5.2 Results

In Fig. 6 we show the cost of uploading FCD in the case of applications that support maximum local fusion, i.e., our best-case aggregation scenario. In the case of RB and RBC, we use a reservation phase of 128 slots. We can notice that all of the distributed algorithms achieve a significant cost reduction, between 75% and 85% at peak hours, and more than 50% even in medium traffic density. The best results are given by RB, with DB being a close contender. However, it must be reminded that DB does not recover information from all the vehicles, which is also why DBC, where FCD from all uncovered vehicles are transmitted individually, gives the worst performance – yet still guaranteeing an important improvement over the non-offloaded baseline scenario. In the case of RBC, we can observe that under medium traffic density, the performance is similar to the one of RB. However, at peak hours, the limited number of slots produces an increased number of collisions, which reduce the algorithms' efficiency. While RB uses the same number of slots in the reservation phase, the number of messages transmitted by RB on the vehicular network is much smaller compared with RBC – $|\mathbb{U}(t)|$ versus $|\mathbb{V}(t)|$ using the notation from Sec. 3.

For applications that do not allow for FCD aggregation, the results are very different, as shown in Fig. 7 :

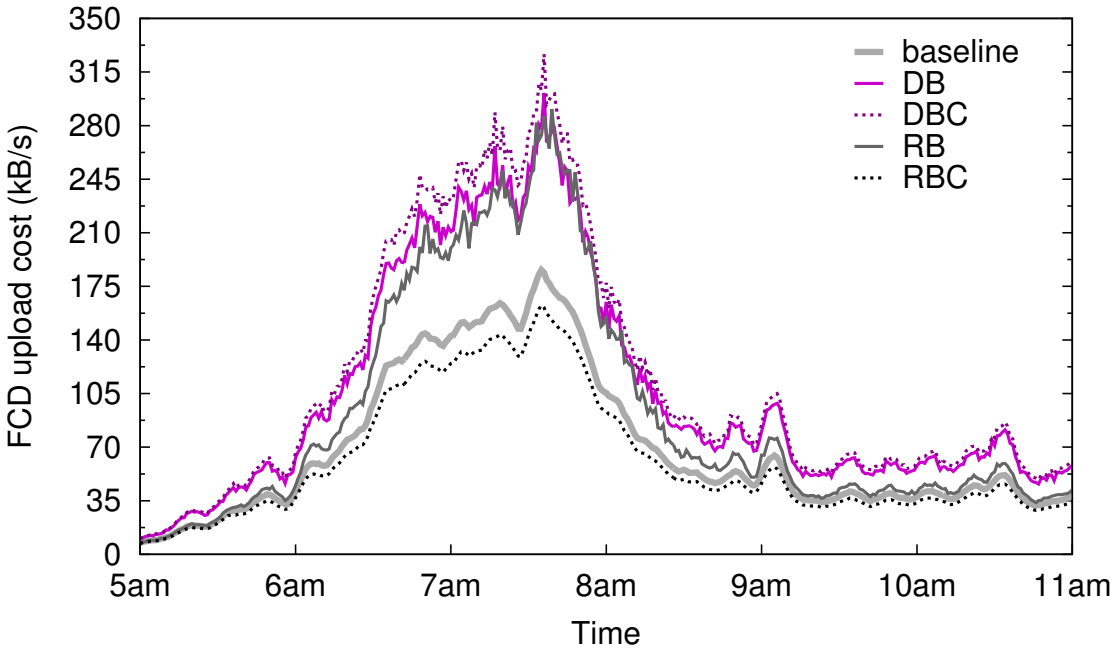


FIGURE 7: Cost of uploading FCD in the worst-case local aggregation scenario using distributed approaches.

DB, DBC and RB incur in a cost that is even higher than that of the baseline case. As discussed in the case of centralized algorithms, this is due to the fact that information coming from nodes with multiple neighboring relays is uploaded more than once. By adding a confirmation phase, RBC manages to solve this problem and reduces the cost by 15-20 kB/s at peak hours compared to a non-collaborative approach.

Similarly to Sec, 4, in Fig. 8 we analyze the impact of the payload size on the performance of the different solutions. Even for small payloads, the three algorithms that do not use any feed-back for the relays (DB, DBC and RB) are less efficient than a non-collaborative approach. On the other hand, RBC with a reservation period of 128 slots always results in a gain, which can be as high as 35% when the payload has a similar size to the TCP/IP headers.

Finally, since from this analysis RBC emerges as a clear winner, we are interested in understanding better its operation. Namely, we investigate the impact of the sole parameters of the algorithm, i.e., the number of slots available in the reservation phase. In Fig. 9, we present results for a reservation phase of different sizes : 8, 32, 64 and 128 slots. We notice that even for a very short reservation phase of 8 slots, RBC still manages a minor cost improvement over the non-offloaded cellular upload. As expected, when the duration of the reservation phase increases, so does the performance of the mechanism. Using more complex reservation slot access techniques, may in fact further reduce the number of collisions and allow RBC to approach the performance of centralized solutions.

Overall, the results presented in this section show that **significant gains can be obtained even by simple distributed algorithms**. For an efficient offloading, **a feed-back for the selected relay nodes is required** in the case when information from individual vehicles must be uploaded. We proposed a solution, RBC, which provides this feed-back with a small overhead of one message per vehicle transmitted on the vehicular network.

6 Conclusion

With an increased number of connected vehicles which produce more and more data, the problem of collecting Floating Car Data in an efficient manner becomes important for end users. In this paper, we study the gain that can be brought by offloading FCD from the cellular network through the use of short-range

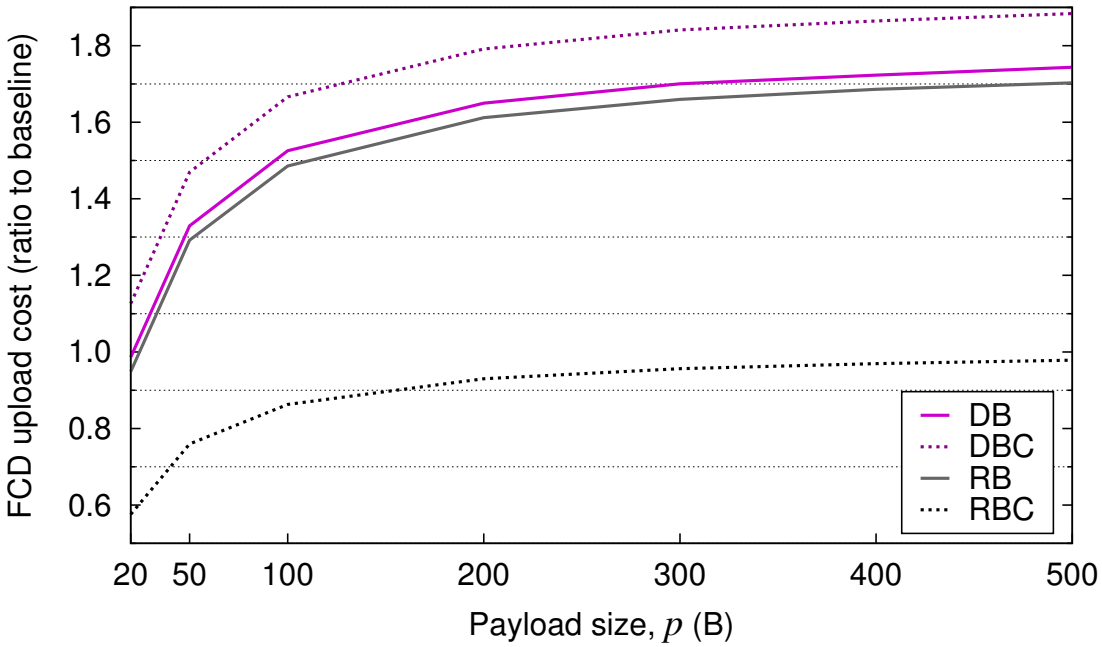


FIGURE 8: Relative cost of uploading FCD (with respect to baseline scenario with no offload) for the worst case aggregation function in the case of distributed algorithms. Results are averaged between 5am and 11am.

vehicle-to-vehicle communication. By considering applications that allow different degrees of in-network local aggregation, **we show that a reduction of more than 40% of the uploading cost is theoretically possible** if the topology of the vehicular network is known. Moreover, **we propose a simple, distributed heuristic that provide effective FCD offload in any aggregation conditions**. Our results indicate that this solution can reach a maximal cost reduction of around 35%, and is generally very close to the optimal centralized performance.

Références

[1] R.N. Charette, *"This car runs on code"*, IEEE Spectrum, Feb. 2009.

[2] U. Paul, A.P. Subramanian, M.M. Buddhikot, S.R. Das, *"Understanding traffic dynamics in cellular data networks"*, IEEE INFOCOM, Shanghai, RPC, Apr. 011.

[3] Cisco, *"Global Mobile Data Traffic Forecast Update, 2013-2018"*, White Paper, Feb. 2014.

[4] R.-P. Schäfer, K.-U. Thiessenhusen, P. Wagner, *"Traffic Information by Means of Real-time Floating Car Data"*, ITS World Congress, Chicago, IL, USA, Oct. 2002.

[5] B.S. Kerner, C. Demir, R.G. Herrtwich, S.L. Klenov, H. Rehborn, M. Aleksić, A. Haug, *"Traffic State Detection with Floating Car Data in Road Networks"*, IEEE ITSC, Vienna, Austria, Sep. 2005.

[6] C. de Fabritiis, R. Ragona, G. Valenti, *"Traffic Estimation And Prediction Based On Real Time Floating Car Data"*, IEEE ITSC, Beijing, PRC, Oct. 2008.

[7] S. Messelodi, C.M. Modena, M. Zanin, F.G.B. De Natale, F. Granelli, E. Betterle, A. Guarise, "Intelligent extended floating car data collection", *Expert System with Applications*, 36(3) :4213–4227, Apr. 2009.

[8] S. Eichler, *"Anonymous and Authenticated Data Provisioning for Floating Car Data Systems"*, IEEE ICCS, Singapore, Oct. 2006.

[9] S. Rass, S. Fuchs, M. Schaffer, K. Kyamakya *"How to Protect Privacy in Floating Car Data Systems"*, ACM VANET, San Francisco, CA, USA, Sep. 2008.

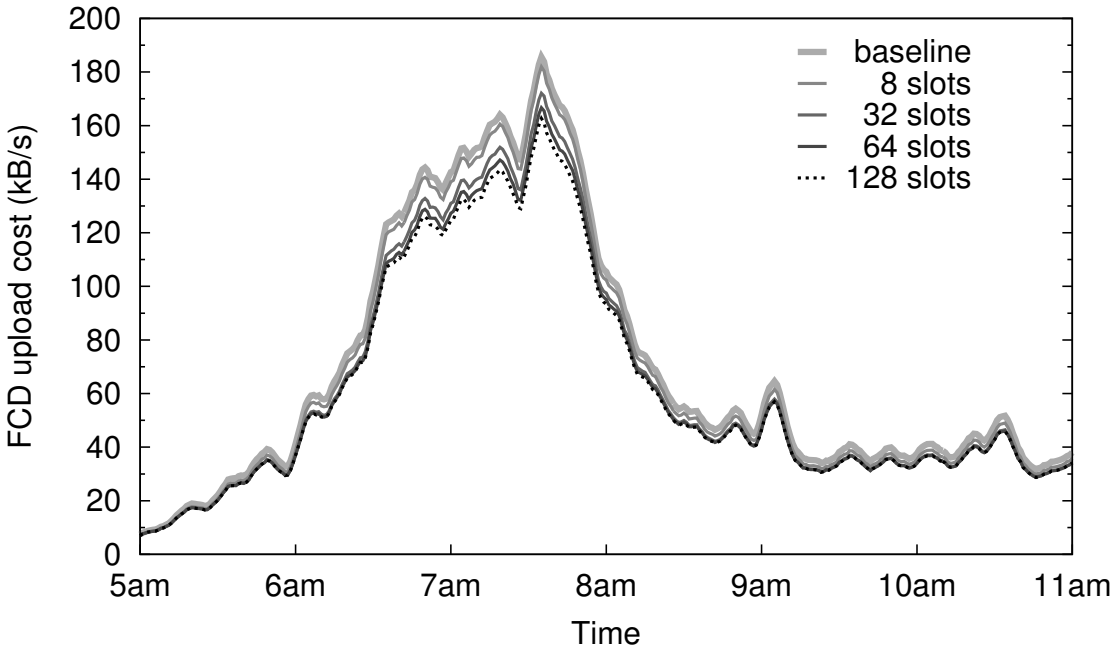


FIGURE 9: RBC algorithm : impact of the number of slots available during each collection period on the FCD upload cost.

[10] C. Sommer, A. Schmidt, Y. Chen, R. German, W. Koch, F. Dressler, *"On the Feasibility of UMTS-based Traffic Information Systems"*, Ad Hoc Networks, 8(5) :506–517, Jul. 2010.

[11] A. Bazzi, B.M. Masini, O. Andrisano, *"On the Frequent Acquisition of Small Data Through RACH in UMTS for ITS Applications"*, IEEE Transactions on Vehicular Technology, 60(7) :2914–2926, Sep. 2011.

[12] C. Ide, B. Dusza, C. Wietfeld, *"Performance Evaluation of V2I-Based Channel Aware Floating Car Data Transmission via LTE"*, IEEE ITSC, Anchorage, AK, USA, Sep. 2012.

[13] T. Mangel, T. Kosch, H. Hartenstein, *"A Comparison of UMTS and LTE for Vehicular Safety Communication at Intersections"*, IEEE VNC, Jersey City, NJ, USA, Dec. 2010.

[14] 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, Jun. 2011.

[15] 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, Sep. 2010.

[16] F. Malandrino, C. Casetti, C.-F. Chiasserini, M. Fiore, *"Offloading Cellular Networks through ITS Content Download"*, IEEE SECON, Seoul, Korea, Jun. 2012.

[17] A. Bazzi, B.M. Masini, G. Pasolini, *"V2V and V2R for Cellular Resources Saving in Vehicular Applications"*, IEEE WCNC, Paris, France, Apr. 2012.

[18] R. Stanica, M. Fiore, F. Malandrino, *"Offloading Floating Car Data"*, IEEE WoWMoM, Madrid, Spain Jun. 2013.

[19] Y. Go, J. Won, D. Kune, E. Jeong, Y. Kim, K. Park, *"Gaining Control of Cellular Traffic Accounting by Spurious TCP Retransmission"*, INET NDSS, San Diego, CA, USA, Feb. 2014.

[20] 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.

[21] 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.

[22] 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.

[23] S. Uppoor, M. Fiore, *"Large-scale Urban Vehicular Mobility for Networking Research"*, IEEE VNC, Amsterdam, Netherlands, Nov. 2011.

[24] A. Larmo, M. Lindstrom, M. Meyer, G. Pelletier, J. Torsner, H. Wiemann, *"The LTE Link-Layer Design"*, IEEE Communications Magazine, 47(4) :52-59, Apr. 2009.