



HAL
open science

SCoD: Saving Cellular Resources by Delaying Transmissions of Popular Content

Salah Eddine Belouanas, Kim-Loan Thai, Prométhée Spathis, Marcelo Dias
de Amorim

► **To cite this version:**

Salah Eddine Belouanas, Kim-Loan Thai, Prométhée Spathis, Marcelo Dias de Amorim. SCoD: Saving Cellular Resources by Delaying Transmissions of Popular Content . IEEE Twelfth International Symposium on Wireless Communication Systems, Aug 2015, Brussels, Belgium. pp.111-115, 10.1109/ISWCS.2015.7454309 . hal-01203370

HAL Id: hal-01203370

<https://hal.science/hal-01203370v1>

Submitted on 1 Oct 2015

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.

SCoD: Saving Cellular Resources by Delaying Transmissions of Popular Content

Salah-Eddine Belouanas, Kim-Loan Thai, Prom  th  e Spathis, and Marcelo Dias de Amorim

LIP6/CNRS – UPMC Sorbonne Universities – Paris, France

{salah-eddine.belouanas, kim.thai, promethee.spathis, marcelo.amorim}@lip6.fr

Abstract—At a time when operators struggle with unprecedented mobile data growth, saving cellular resources is of utmost importance. The first traffic that we must get rid of is redundant traffic. This happens, for example, when popular data must be disseminated to a population of subscribers. We propose SCoD, a strategy that benefits from the delay-tolerance of some applications to defer transmissions to appropriate instants. The idea is to wait for nodes to gather around a minimum number of access points so that the total number of transmissions is reduced. SCoD leverages as much as possible node mobility for the sake of clustering; to this end, SCoD relies on different guiding functions that trigger, whenever appropriate, a transmission. We assess the performance of our strategy by running simulations on a realistic vehicular dataset and find out that it leads to significant improvements when compared with traditional (instantaneous) dissemination strategies.

I. INTRODUCTION

Cellular operators are struggling with the skyrocketing growth of mobile traffic these latest years [1]. In a near future, they will likely have no choice but to combine a wide range of techniques to save cellular resources. So far, the main efforts have focused on increasing the cellular capacity, by either adopting new standards (e.g., 3G→4G→5G), or deploying more antennas, or acquiring new licensed bands.

At the same time, operators also adopt other strategies to limit usage of cellular resources. For example, most operators implement a two-phase resource assignment strategy, where users see their bandwidth capped after some data consumption threshold (generally in the order of a few Gbytes) [2], [3]. A more recent strategy consists in offloading traffic through either an alternative technology (e.g., from cellular to Wi-Fi), or in-band channels (e.g., using small cells) [4], or even device-to-device communications [5]. Following such recent trends, we are convinced that *any* saving is welcome.

In this paper, we propose to reduce the consumption of wireless resources by better controlling the number of transmissions, whenever possible. In particular, we advocate that a simple policy such as *when* to transmit could lead to significant savings in situations where data is popular and tolerates some delay. Fortunately, these situations are not rare.

The infrastructure we consider consists of a set \mathbf{A} of access points covering a population \mathbf{N} of mobile users interested in a content c . In this paper, unless specified, we do not deal with any specific radio technology and will simply refer to $a_i \in \mathbf{A}$ as a wireless access point (or base station) that covers

a zone called cell. Full dissemination must be achieved before a deadline τ . Our goal is to determine, on the fly, the minimum subset of access points $\hat{\mathbf{A}} \subset \mathbf{A}$ that should transmit a copy of the content to cover the entire population \mathbf{N} . The challenge is to compute the different instants when access points should transmit the content; this is difficult, as user mobility cannot be known in advance.

To address the abovementioned problem, we propose SCoD (Scheduled Content Delivery), a strategy to disseminate delay-tolerant contents by planning their transmission whenever groups of users in the same cell are considered big enough to trigger a transmission. Deciding what “big enough” means is the role of SCoD. We assume that a single transmission by an access point is received by all users in the corresponding cell. We also assume that each access point knows the list of mobile users it covers at all times. The population of users is composed of two types of users: *sane users* (who have not yet received the content) and *infected users* (who have already obtained the content). The decision to transmit or not is taken based on a *decision function* (DF), whose goal is to determine the threshold above which a group of users is sufficiently big to trigger a transmission. In a nutshell, SCoD relies on decision functions whose *aggressiveness* decreases over time (i.e., the threshold to consider that a group is big enough is higher in the beginning of the period). At the end of the process, when the content reaches the deadline, a content copy is transmitted within every cell hosting at least one sane user, so that full coverage is guaranteed.

We illustrate the idea of SCoD in Fig. 1. By tracking where sane users are, the infrastructure can figure out which cells host a sufficiently large number of sane users to determine if the corresponding access point should transmit the content. This is to be compared with a baseline strategy (which does not spread transmissions over time) that would require, most of the time, more transmissions in the beginning of the period (we will see this kind of behavior in Section IV). We assess the performance of SCoD by investigating several decision functions. Simulation results were obtained for a realistic large-scale vehicular trace of Bologna city (Italy). We observe that SCoD significantly reduces the number of transmissions when compared with the baseline approach. Furthermore, SCoD is very close to an oracle solution (unpractical, as it assumes the knowledge of future user displacements). Moreover, the number of satisfied users converges very quickly to the total population.

The remainder of this paper is organized as follows. In

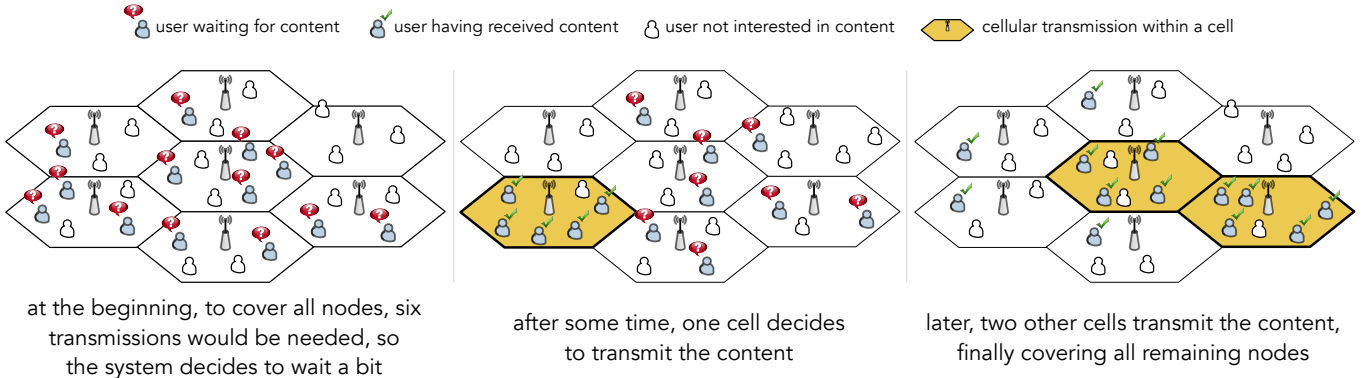


Fig. 1: Instead of transmitting six copies of the content at the beginning, the idea behind SCoD is to determine the right instants to trigger a transmission. In the example of the figure, it is possible to cover all users with only three transmissions. Although for a single content it might be negligible, accumulated savings can be huge on the long term.

Section II, we present the system model and state the problem we consider in this paper. We introduce SCoD in Section III and motivate the role of the different decision functions. In Section IV, we present the evaluation methodology and the results. We postpone some related work to Section V and conclude the paper in Section VI.

II. CONTEXT AND PROBLEM STATEMENT

Without loss of generality, we model our system as a cellular network with limited capacity on the wireless access channel. We consider a system defined by the tuple $[c, \mathbf{N}, \mathbf{A}, t_0, \tau]$, where c is the content to be disseminated, \mathbf{N} is the set of users that are interested in receiving c , \mathbf{A} is the set of access points, t_0 is the time when c becomes available at the infrastructure, and τ is the maximum delay that any user tolerates before receiving c . We assume that c is generated at t_0 . At a given time $t \in [t_0, t_0 + \tau]$, the set of users are organized into two classes: $\mathbf{N}_{a_i}^{\text{sane}}(t)$ (the set of sane users associated with access point $a_i \in \mathbf{A}$ at time t) and $\mathbf{N}_{a_i}^{\text{inf}}(t)$ (the set of infected users associated with access point $a_i \in \mathbf{A}$ at time t). At t_0 , the distribution of users is as follows: $\mathbf{N}_{a_i}^{\text{sane}}(t_0) = \mathbf{N}$ and $\mathbf{N}_{a_i}^{\text{inf}}(t_0) = \emptyset$, for all $a_i \in \mathbf{A}$.

All cells are supposed to be disjoint. Thus, each (mobile) user is associated with one and only one access point at all times. For the purposes of this paper, we also assume that the population \mathbf{N} remains unchanged during $[t_0, t_0 + \tau]$ (i.e., churn is not considered).

As previously mentioned, content c becomes available at t_0 and must be delivered to sane users by $t_0 + \tau$. We assume perfect transmissions, i.e., a single transmission is enough to reach all users in the cell. Whenever an access point decides to transmit c , all sane users under this access point become infected users. Our goal consists in minimizing the number of transmissions by leveraging the mobility of the users. We simplify the problem by discretizing the content lifetime in time steps, then we examine each cell independently of the others. As time goes by, users may move from cell to cell; thus, density within a cell changes over time. We call a given distribution of users throughout the cells a *configuration*.

Problem definition. Given the tuple $[c, \mathbf{N}, \mathbf{A}, t_0, \tau]$, find a strategy to determine a threshold for the number of users in a cell at some time $t \in [t_0, t_0 + \tau]$ so that the corresponding access point decides to transmit c .

Note that setting a small threshold may lead to too many transmissions, while a big threshold may refrain an access point from transmitting in appropriate occasions. Finding a good trade-off is challenging, especially because the mobility of the users cannot be known in advance.

III. SCoD: SPECIFICATION AND OPERATION

We design SCoD following the different issues and assumptions discussed in the previous section. The most important features of SCoD are illustrated in Fig. 2. At t_0 , content c becomes available and must be delivered to all target subscribers by $t_0 + \tau$. Upon each configuration change, SCoD checks whether the content deadline has been reached or not; if so, the decision is to transmit c in every cell containing at least one sane user. If not, a specific function $\beta(t)$ (called *lateness factor*) is used to decide whether an access point should transmit or not. Function $\beta(t)$ changes over time as it determines the *aggressiveness* of the decision process. SCoD is more aggressive in the beginning of the period, i.e., it expects bigger groups of users in a cell. As time goes by, it relaxes by tolerating smaller groups, up to the deadline where all cell containing at least a sane user must transmit. This procedure is formalized as follows. Access point a_i transmits c at time t if one of the two conditions hold:

$$|\mathbf{N}_{a_i}^{\text{sane}}(t)| \geq \beta(t) \sum_{j=1}^A |\mathbf{N}_{a_j}^{\text{sane}}(t)| \ \& \ \beta(t) \neq 0 \quad \text{OR} \quad (1)$$

$$|\mathbf{N}_{a_i}^{\text{sane}}(t)| > 0 \ \& \ \beta(t) = 0$$

where $|\mathbf{N}_{a_i}^{\text{sane}}(t)|$ is the number of sane users associated with a_i at time t and $\beta(t)$, the lateness factor is a time-based function ($0 \leq \beta(t) \leq 1$) whose main role is to delay the transmissions. In other terms, it defines whether the system should be conservative and keep the content longer. We detail this function hereafter.

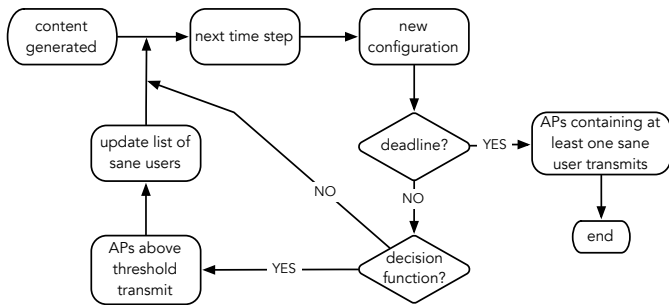


Fig. 2: High level operation of SCoD.

Whenever a transmission takes place in at least one cell (note that nothing prevents multiple access points from transmitting at the same time), the system updates the list of sane users before the next configuration. The system carries on at the same pace until the deadline is reached. However, while the time is running, β decreases, which causes, at a given time, one of the inequalities in (1) to be satisfied. As a consequence, one or multiple transmissions occur.

SCoD transmission decision functions. We tested different functions as possible expressions of $\beta(t)$. As we want the aggressiveness to decrease over time, we pick functions that are monotonically decreasing. These functions are listed in Table I.

TABLE I: Lateness factor functions used for $\beta(t)$.

Function	Expression
Logarithmic	$\beta(t) = \ln\left(\frac{-t+\tau}{\tau} + 1\right)$
Linear	$\beta(t) = \frac{-t}{\tau} + 1$
Square-root	$\beta(t) = \sqrt{\frac{-t+\tau}{t+\tau}}$
Exponential	$\beta(t) = 1 - e^{\frac{t}{\tau}-1}$

IV. EVALUATION

A. Evaluation setup

In order to evaluate SCoD, we use the Bologna vehicular dataset, which contains a significant number of mobile nodes [6]. The dataset covers 20.6 km^2 in the city center for a period of two hours, featuring the movement of 10333 vehicles. We compare SCoD with three reference strategies: full coverage, immediate transmission, and oracle. In *full coverage*, the infrastructure systematically distributes the content in every cell (even if the cell does not have any user). Of course, this strategy is too naive to be implemented, and serves only as a lower bound. In *immediate transmission*, the content is transmitted at t_0 in all cells hosting sane users. Finally, in *oracle*, the decision to transmit is based on the prior knowledge of all future configurations during the whole period. This strategy is non-causal (thus not realizable), but provides an upper bound for the expected performance. The previous three strategies are simulated for a starting value of the lateness factor of $\beta(t = t_0) = 0.25$. The value of β at $t = t_0$ gives the initial aggressiveness of SCoD. For example,

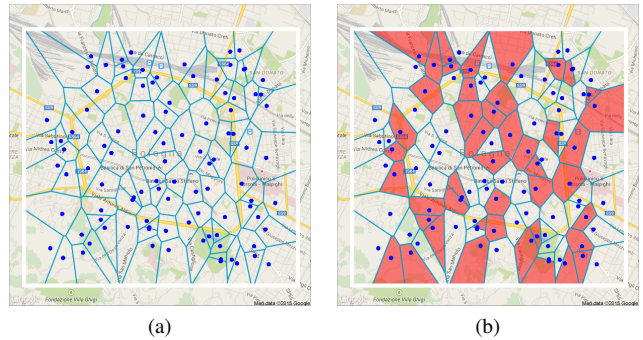


Fig. 3: (a): The Voronoi diagram for a set of 100 APs deployed in the city of Bologna. (b) The red cells are the one involved in the transmission process of a content.

when $\beta(t = t_0) = 0.25$, it means that 25% of the nodes must be associated with an access point to trigger a transmission.¹

We focus on a specific area of the city of Bologna where we track the location of each vehicle for a period of two hours. To assign each vehicle to an access point, we emulate a deployment plan of a wireless infrastructure consisting of access points composing the set \mathbf{A} . Each vehicle is assigned to the nearest access point by using a Voronoi diagram [7]. Each cell of the Voronoi diagram contains exactly one access point. We depict in Fig. 3a the tessellation for 100 access points within the central area of Bologna.

To give more confidence to our results, we iterate the process of partitioning the map of Bologna by emulating multiple random deployment plans following a uniform distribution. The various topologies are generated for 100 and 300 access points. We assume that a single transmission by an access point is received by all vehicles in the corresponding cell. The results we show are provided for 1,000 and 2,000 nodes (vehicles in our scenario). With regard to the content deadline, we consider two scenarios with the following values: 300 s and 1,200 s (realistic values when considering such content as: pollution indication, important event announcements, popular content in social media etc). The evaluation performance results we present in the next section reflect averages over 20 simulation runs for different generated topologies (spatial distribution of access points). We primarily focus on the *average transmission gain* of a given strategy, $G_{\text{strategy}}^{\text{avg}}$, defined as:

$$G_{\text{strategy}}^{\text{avg}} = \frac{N_{\text{FC}} - N_{\text{strategy}}}{N_{\text{FC}}}, \quad (2)$$

where, N_{FC} is the number of transmissions when applying the full coverage strategy, and N_{strategy} is the number of transmissions of the evaluated strategy (oracle, immediate transmission, and SCoD). Another performance criterion is the *user satisfaction* degree, which refers to how fast a strategy satisfies all the users at the lowest possible cost.

B. Evaluation results

We will show that SCoD performs very well in saving cellular resources by decreasing the number of transmissions.

¹This value is quite aggressive, but still leads to good results as we will show later in Section IV-B.

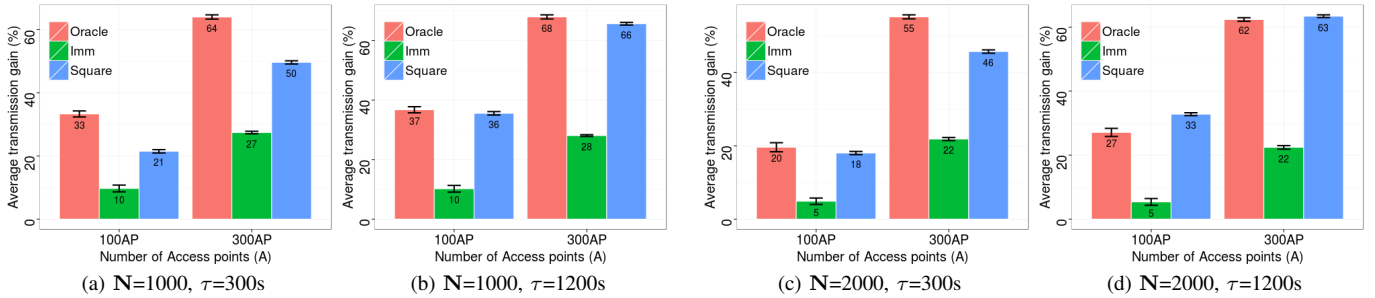


Fig. 4: Average transmission gain for SCoD (using a “square” lateness function) and two baseline strategies (Oracle, Immediate). Different values of the content lifetime and the users population are presented. Confidence intervals of 95% are shown as well.

In the following, we will show the results for only one lateness function (we pick the “square” function). As a matter of fact, we observed that all functions described in Table I led to results with a small differences. In other words, this means that the aggressiveness aspect is more important than the skewness of the curve. We will not delve into this point in this paper, and will leave this discussion for future work.

Average transmission gain. We consider “full coverage” as the baseline reference to assess the gains obtained by each of other strategies. We use the *average transmission gain* $G_{\text{strategy}}^{\text{avg}}$ as defined in Section IV-A. We plot in Fig. 4 the average transmission gain for multiple combinations of N and τ . As expected, the longer the content lifetime, the more SCoD gains. This is inherently related to the philosophy behind SCoD, whose goal is to benefit as much as possible from the mobility of the nodes. An interesting observation is that we achieve almost the same performance level as the oracle, especially for longer τ . The main reason for this behavior is that users have more time to wave among cells, which allows SCoD having more freedom to select the most appropriate time to transmit the content.

We observe that the population size also has a positive impact on the gains. In one particular case, SCoD even outperformed the oracle. Although the oracle has the advantage of knowing the whole mobility trace in advance, it is not optimal. In fact, implementing the oracle is equivalent to determining the minimum cover set of a graph, which is known to be a NP-complete problem [8]. In this paper, we used the greedy strategy as the heuristic to compute the oracle, which gives some room for improvements (which happened, for example, in Fig. 4d) [9].

User satisfaction. We are also interested in evaluating the evolution of the dissemination process. In Fig. 5, we plot the CDF of satisfied users in function of the transmissions. There are a few observations we can make here. First, we see again that SCoD covers the same amount of nodes as the immediate strategy, but with less transmissions. SCoD’s and oracle’s curves overlap when the content tolerates longer delays. Second, and more interestingly, SCoD shows a much more steep slope. This means that SCoD limits the number of transmissions in lowly-populated cells. As a consequence, SCoD is more robust to transmission failures when the dead-

line approaches (failures that happen earlier are easier to recover as we have more time left).

V. RELATED WORK

The problem of disseminating content while alleviating the burden on the overloaded infrastructure has been well studied in the literature. First proposals in this sens have target at better understanding node mobility and delay tolerance of some types of content. Solazzo et al. propose TACO-DTN, a content dissemination system that uses the info-station model to manage the resources of an access point when disseminating information [10], [11]. In the area of vehicular networking, similar approaches have been considered. Luan et al. propose a practical and cost-effective approach to construct a fully distributed roadside communication infrastructure to facilitate localized content dissemination to vehicles in a urban scenario [12]. Also in a vehicular context, Ahn et al. leverage both infrastructure and peer-to-peer communications to save resources at the cellular infrastructure [13].

Some effort have been done to maximize content dissemination within a deadline, while minimizing the use of cellular resources. Most works fall under the umbrella of *mobile data offloading*, which is considered to be an inexpensive method to improve the capacity of mobile networks. Balasubramanian et al. were the first to propose to delay transmissions until a node reaches a Wi-Fi hotspot [14], whose cost is assumed to be negligible when compared with the one of the cellular infrastructure. Similarly, Lee et al. suggests significant savings in energy and traffic by also delaying traffic [15]. To this end, they perform extensive analyses of traces of pedestrian mobility collected in a real metropolitan environment. From a much more theoretical perspective, Mehmeti and Spyropoulos propose a queueing analytic model for delayed offloading, and derive the mean delay, offloading efficiency, and other metrics of interest, as a function of the delay tolerance as well as other network parameters [16]. Other offloading techniques rely on device-to-device communications. Rebecchi et al. [17], Whitbeck et al. [18], and Han et al. [19] have considered scenarios in which device-to-device and cellular connections are used to disseminate the content, considering the social ties and the geographical proximity.

More related to our work, an alternative approach has been the center of focus in current literature, depends on the in-

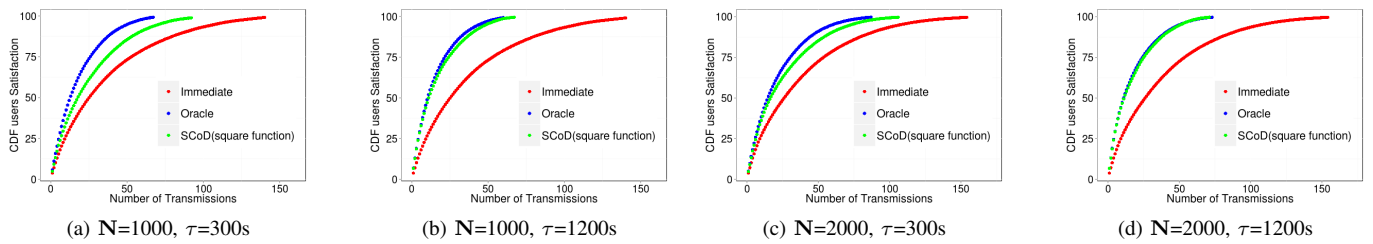


Fig. 5: CDF of the users satisfaction for each strategy vs the number of transmission and the gain, and $\mathcal{A}=300$ for all the plots.

stantaneous availability of resources to deliver content without wasting resources [20]. Although these contributions do show interesting results, they assume that nodes can communicate in ad hoc mode, which is not the case of this paper.

To summarize, our proposal differs from the previous works as it deals with full dissemination guarantees under a very constraint capacity of a cellular infrastructure.

VI. SUMMARY AND OUTLOOK

We proposed SCoD, a new content dissemination strategy specifically designed for cellular networks that suffer from heavy capacity constraints at the wireless access channel. It leverages node mobility and delay-tolerance to determine the best possible instants at which an access point must transmit a copy of a given content to a multitude of receivers. In a nutshell, its goal is to let users gather within fewer cells to reduce the number of transmissions. Experiments using a realistic vehicular dataset in the city of Bologna reveal that SCoD manages to reduce the number of transmissions when compared with strategies that do not play with the mobility of the nodes, while achieving 100% content delivery.

Our work will continue in the following directions. Firstly, we will address the problem of handling multiple contents that achieve a deadline at the same time. This will require the proposal of prioritization strategies to balance the dissemination evolution of the different contents. Secondly, we will investigate churn, where the system would accept new arrivals and departures. Thirdly, we will investigate the sensibility of the system with regard to the lateness function and its initial value.

ACKNOWLEDGMENT

This work was partially funded by the French National Research Agency (ANR) under project ANR DataTweet (ANR-13-INFR-0008).

REFERENCES

- [1] Cisco, "Cisco visual networking index: Forecast and methodology, 2013-2018," 2014.
- [2] At&t accused of deceiving smartphone customers with unlimited data plans. [Online]. Available: http://www.nytimes.com/2014/10/29/technology/FTC-says-ATT-deceived-consumers-on-unlimited-data-plan.html?_r=0
- [3] Mobile internet: How many gigabytes do you need ? [Online]. Available: <http://kenstechtips.com/index.php/what-does-500mb-or-1gb-internet-actually-mean-explaining-mobile-data-limits>
- [4] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, "Data offloading techniques in cellular networks: A survey," 2014.
- [5] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," 2014.
- [6] itetris project. [Online]. Available: <http://www.ict-itetris.eu/>
- [7] F. Aurenhammer, "Voronoi diagrams: a survey of a fundamental geometric data structure," *ACM Computing Surveys (CSUR)*, vol. 23, no. 3, pp. 345–405, 1991.
- [8] R. M. Karp, "Reducibility among combinatorial problems," *Complexity of Computer Computations*, 1972.
- [9] V. Chvatal, "A greedy heuristic for the set-covering problem," *Mathematics of Operations Research*, vol. 4, no. 3, pp. 233–235, 1979.
- [10] G. Sollazzo, M. Musolesi, and C. Mascolo, "Taco-dtn: a time-aware content-based dissemination system for delay tolerant networks," in *Proceedings of the 1st international MobiSys workshop on Mobile opportunistic networking*. ACM, 2007, pp. 83–90.
- [11] R. H. Frenkiel, B. Badrinath, J. Borras, and R. D. Yates, "The info-stations challenge: Balancing cost and ubiquity in delivering wireless data," *IEEE Personal Communications*, vol. 7, no. 2, pp. 66–71, 2000.
- [12] T. H. Luan, L. X. Cai, J. Chen, X. Shen, and F. Bai, "Engineering a distributed infrastructure for large-scale cost-effective content dissemination over urban vehicular networks," *Vehicular Technology, IEEE Transactions on*, vol. 63, no. 3, pp. 1419–1435, 2014.
- [13] J. Ahn, M. Sathiamoorthy, B. Krishnamachari, F. Bai, and L. Zhang, "Optimizing content dissemination in vehicular networks with radio heterogeneity," *Mobile Computing, IEEE Transactions on*, vol. 13, no. 6, pp. 1312–1325, 2014.
- [14] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3g using wifi," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM, 2010, pp. 209–222.
- [15] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile data offloading: how much can wifi deliver?" *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, Apr. 2013.
- [16] F. Mehmeti and T. Spyropoulos, "Is it worth to be patient? analysis and optimization of delayed mobile data offloading," in *IEEE Infocom*, Toronto, Canada, May 2014.
- [17] F. Rebecchi, M. Dias de Amorim, and V. Conan, "Droid: Adapting to individual mobility pays off in mobile data offloading," in *Networking Conference, 2014 IFIP*. IEEE, 2014, pp. 1–9.
- [18] J. Whitbeck, M. Amorim, Y. Lopez, J. Leguay, and V. Conan, "Relieving the wireless infrastructure: When opportunistic networks meet guaranteed delays," in *World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2011 IEEE International Symposium on a*. IEEE, 2011, pp. 1–10.
- [19] B. Han, P. Hui, V. Kumar, M. V. Marathe, G. Pei, and A. Srinivasan, "Cellular traffic offloading through opportunistic communications: a case study," in *Proceedings of the 5th ACM workshop on Challenged networks*. ACM, 2010, pp. 31–38.
- [20] P. Lungaro, Z. Segall, and J. Zander, "Context-aware rrm for opportunistic content delivery in cellular networks," in *Communication Theory, Reliability, and Quality of Service (CTRQ), 2010 Third International Conference on*. IEEE, 2010, pp. 175–180.