Quantifying the Potential of Ride-Sharing using Call Description Records

Blerim Cici[†⋆], Athina Markopoulou[†], Enrique Frías-Martínez[⋆], Nikolaos Laoutaris[⋆]
UC Irvine[†], Telefonica Research(Spain)[⋆]
{bcici, athina}@uci.edu, {efm, nikos}@tid.es

ABSTRACT

Ride-sharing on the daily home-work-home commute can help individuals save on gasoline and other car-related costs, while at the same time reducing traffic and pollution in the city. Work in this area has typically focused on technology, usability, security, and legal issues. However, the success of any ride-sharing technology relies on the implicit assumption that human mobility patterns and city layouts exhibit enough route overlap to allow for ride-sharing on the first place. In this paper we validate this assumption using mobility data extracted from city-wide Call Description Records (CDRs) from the city of Madrid. We derive an upper bound on the effectiveness of ride-sharing by making the simplifying assumption that any commuter can share a ride with any other as long as their routes overlap. We show that simple ride-sharing among people having neighboring home and work locations can reduce the number of cars in the city at the expense of a relatively short detour to pick up/drop off passengers; e.g., for a 0.6 km detour, there is a 52% reduction in the number of cars. Smartphone technology enables additional passengers to be picked up along the way, which can further reduce the number of cars, as much as 67%.

1. INTRODUCTION

Ride-sharing is an effective way to reduce the number of cars on the streets in order to address both individual and city-wide issues. On one hand, individuals are interested in reducing the cost of their car usage and save on gasoline and other usage-based costs [2]. On the other hand, cities are interested in reducing traffic and pollution and provide incentives (e.g. reserved carpooling lanes) to encourage commuters to share rides. In recent years, a plethora of web- or smartphone-based solutions have emerged in order to facilitate intelligent traffic management [18], [17], [5], and in particular ride-sharing. Systems like `carpooling.com`, and `eRideShare.com` have attracted a few million users in Europe and the US but the technology hasn't been widely adopted yet.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
ACM HotMobile'13, February 26–27, 2013, Jekyll Island, Georgia, USA.
Copyright 2013 ACM 978-1-4503-1421-3 ...$10.00.

Ride-sharing systems started in the US during WW-II. These early "word-of-mouth" systems required predefined rendezvous and prior familiarity among commuters, which limited the number of neighbors a person could share a ride with. More recently, web-based solutions, such as `Amovens.com`, allow drivers and passengers to advertise their interest in ride-sharing, thereby increasing the chances of finding a match, but still generally require predefined rendezvous. Using smartphones with ride-sharing apps, such as `Avego.com`, allows drivers and passengers to be matched opportunistically without pre-arranged rendezvous. Such systems are promising but it is still unclear whether they will be widely adopted.

Work in the area has focused on technological, usability, security and legal aspects [16] [20]. Previous research has shown that ride-sharing has economic advantage over driving alone, and that is more spatially flexible and less time consuming than public transportation, but it is not sure if this advantages are strong enough to entice commuters to switch to ride-sharing; privacy and flexibility are major reasons why the vast majority of commuters choose to drive alone. Many believe that current technology provides insufficient levels of security and monitoring to allow people to travel safely with strangers; others believe that it is only an unsolved bootstrapping problem that keeps the technology from booming. Most people, however, implicitly assume that human mobility patterns and the layouts of today's cities exhibit enough route overlap for ride-sharing to take off, once the aforementioned issues are solved.

In this paper, we validate this underlying assumption, which is crucial for the success of any ride-sharing system. To this end, we use home/work locations, for a large population of a city, extracted from a CDR dataset; the inferred home and work locations are used to match people in groups that share the same car. The exact potential of ride-sharing depends on many factors, not all of which can be known at the scale of our study (e.g., behavioral traits). Our approach is to focus only on quantifiable factors and mask all other unknown factors by making the simplifying assumption that ride-sharing is constrained only by the distance of end-points and time. Therefore, our quantification is actually an upper bound of the exact potential of ride- sharing that may be constrained by additional factors.

Our contributions are as follows. We consider two scenarios for ride sharing, and for each scenario we develop an efficient algorithm to do the matching, and quantify the benefit of ride-sharing in terms of reduction in the number of cars. Note that, the theoretical limit of car reduction is 75%

(all users are matched in cars of 4 and only 25% of the cars are used). The scenarios and results are summarized below.

`EndPoints RS`: In this scenario, we consider ride-sharing only among users with nearby (i.e., within distance δ) home and work locations. We formulate the problem of matching users so as to minimize the number of cars, and serve all rides among neighbors that are within distance δ in both home and work location, as a *Capacitated Facility Location Problem with Unsplittable Demand*. Since the latter is an NP-hard problem, we develop an efficient heuristic algorithm that we name `EndPoints RS`. When applied to our CDR dataset, the algorithm provides an upper bound to the ride-sharing potential benefits: 52% car reduction for δ to 0.6 km, if we assume that drivers wait for passengers as long as necessary. Although unrealistic, this puts an upper bound on the savings of ride-sharing based only on spatial information about home and work. When we introduce time constraints into the picture, we find that the effectiveness of ride-sharing becomes proportional to the driver/passenger waiting time for a pick-up, and inversely proportional to the standard deviation of departure times. For example, with a standard deviation of 10 minutes, a wait time up to 10 minutes, and a δ of 0.4 km there is a 26% reduction of cars.

`EnRoute RS`: Next, we compute the routes from home to work using Google Maps and allow en-route pick-ups. First, we match neighbors using the `EndPoints RS` algorithm. Then additional passengers are picked up along the way, which clearly increases the effectiveness of ride sharing, using an iterative algorithm, which we refer to `EnRoute RS`. The same δ and τ applied to end points are applied to en-route pick-ups too. For example, with a δ of 0.6 km the effectiveness on savings jumps from 52% for `EndPoints RS` to 67% for `EnRoute RS`. Taking also into account the randomness in departure times, the corresponding numbers are 35% for `EndPoints RS` and 56% for `EnRoute RS`.

The rest of this paper is organized as follows. In Section II we briefly describe the CDR dataset that is the basis of this study. In Section III, we present our methodology for inferring the Home Work location for individual users. In Section IV and Section V we consider matching based on end points (`EndPoints RS`) and en route (`EnRoute RS`) respectively. In particular, we present efficient matching algorithms and we evaluate the benefits of ride-sharing when applying these algorithms on the CDR dataset. Section VI discusses related work. Section VII concludes the paper.

2. OUR CDR DATASET

A valuable asset of this study is the Call Description Records or CDR Dataset, which we obtained from a large cellphone provider. It contains 820M calls from 4.7M mobile users, during a 3-month period, in the metropolitan area of Madrid, Spain.

CDRs are generated when a mobile phone makes or receives a phone call or uses a service (e.g., SMS, MMS, etc.). Information regarding the time/date and the location of the nearest cell tower are then recorded. More specifically, the main fields of each CDR entries are the following: (1) the originating number (2) the destination number (3) the time and date of the call (4) the duration of the call (5) the latitude and longitude of the cell tower used by one, or both, phones numbers — cell phone companies save CDR records only for their customers. These records are logged for pricing purposes, so they come at no extra cost to the cellular operator. Note that no information about the exact position of a user is known, since cell phone data provide coarse location accuracy — a few hundred meters for city center, and up to 3 km in rural areas. For privacy reasons, no contract or demographic data were made available to us, and the originating and destination phone numbers were anonymized. More details about CDRs can be found in [15].

(a) Headquarters of Telefonica in Madrid

(b) Residential Area: Latitude:$40^o30'13.45''$, Longitude:$3^o38'07.69''$

Figure 1: Example of strictly residential and strictly working areas

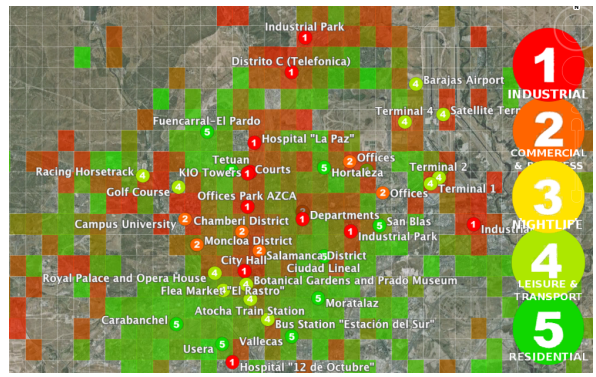


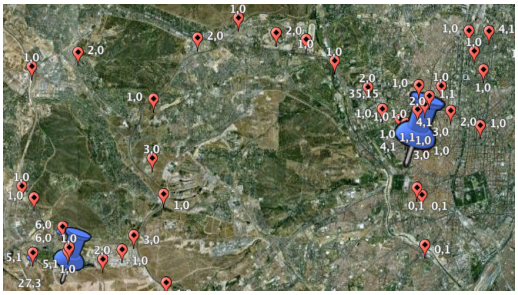
Figure 2: Characterizing Madrid based on our results

3. INFERRING HOME AND WORK

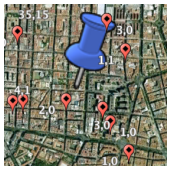
We use an existing methodology to infer the Home/Work locations of 270K individuals in our CDR dataset. This subset of individuals is then used as input to the matching algorithms, discussed in the following section, and the benefit of ride-sharing (reduction in the number of cars) is computed.

3.1 Methodology

We apply the methodology of Isaacman et al. [10], which identifies important places (*i.e.*, places that the user visits frequently and spends a lot of time) in the life of mobile phone subscribers, based on (i) CDR data and (ii) ground truth for a subset of subscribers. More specifically, in [10], the recorded cell towers of a user are clustered to produce the list of places that she visits. Then, regression analysis is applied on the ground truth users (their identified clusters and their true important locations) to determine the features that distinguish clusters that represent important places. Such features include: (1) the number of days that the user appeared on the cluster (2) the duration of user appearances on the cluster (3) the rank of the cluster based on number of days appeared. Once important locations have



(a) A ground truth user



(b) Zooming at home

(c) Zooming at work

Figure 3: The red marks show the recorded cell towers for the user, while the blue marks the clusters. The white numbers next to each mark indicate the number of weekdays and the number of weekends the user appeared in that location. Also, the size of each mark is proportional to the number of days the user has appeared in that location.

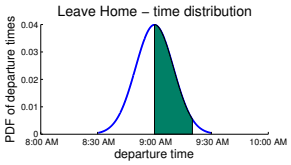
been identified, the authors choose which of these locations is home and which is work. The best features to make this decision are: (4) number of phone calls between 7PM - 7AM (5) number of phone calls between 1PM - 5PM, referred to as *Home Hour Events* and *Work Hour Events* respectively.

In this paper, first, we filter out users for whom we simply do not have enough data: *i.e.* users with less than 1 call per day on average, or less than 2 clusters with 3 days of appearance and 2 weeks of duration. Then, we tune the methodology of [10] in our case. More specifically, we build two classifiers, one for home and one for work, and we train them using the 5 features described above and the ground truth described in 3.2. Once the training is complete, we apply the classifiers to the rest of the users. This was necessary since our ground truth contains only home and work location, while the ground truth of [10] contained other important locations too. Finally, after classification, we keep only the users who have only one location identified as home, and a different location identified as work, since we are interested only in commuters.

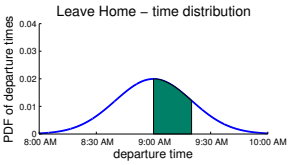
3.2 Obtaining the Ground Truth

In [10], a small set of 37 volunteers who reported their most important locations, including home and work. This information was used to train the classifiers that were applied to the remaining dataset of around 170K mobile phone users.

In our case, we did not have access to the actual phone numbers. We obtained our ground truth for a select subset of users based on our knowledge of the city of Madrid. In particular, due to its development pattern in the last 20 years, Madrid has many areas lying around its outer ring highways that are strictly residential and other ones that are strictly industrial. An example of the former are large



(a) When $\sigma = 10$



(b) When $\sigma = 20$

Figure 4: Assuming users u leaves home at 9:10, the users departing with 10 min difference are in the green area under the curve.

residential development projects in previously isolated areas like the one depicted in Fig. 1. Such areas offer a clear distinction between home and work and can be exploited to build our "ground truth". To this end, we selected 160 users that appeared for many days in only one such residential area during 7PM - 7AM (assumed to be "home" hours), and only one such industrial area during 1PM - 5PM (assumed to be "work" hours). Then, the location inside the residential area is pointed as the user's Home, while the location inside the industrial area is pointed as the user's work.

For each one of the 160 users, we visually inspected their recorded locations through Google Earth for a sanity check. In Fig. 3 we show one of these users: this person lives in the location shown in Fig. 3(b), which is the top right blue marker of Fig. 3(a), and work on the location shown in Fig. 3(c), which is the bottom left marker of Fig. 3(a).

3.3 Results

Applying the above methodology to our CDR dataset, we are able to infer the home and work locations of 270K individual users.

The following comparison provides a sanity check. In Fig. 2 we break the city of Madrid into a grid and paint each square of the grid with a combination of green and red. If the number of inferred home locations is higher than the number of work locations, then the color of the square is closer to green, otherwise it is closer to red. We use an existing study [15] to obtain a characterization of locations in Madrid (industrial, commercial, nightlife, leisure and transport, residential). We annotate such areas in Fig.2 using numbered circles, *e.g.*, the headquarters of Telefonica in Madrid is one of the two red circles on the top of the figure. We observe that the squares that we painted red contain more circles indicating industrial and commercial zones, than residential zones. Also, squares painted green contain more residential zones than industrial zones.

However, due to privacy reason, validating home/work remains challenging. We are currently in the process of obtaining permission to compare our results with the billing database of the operator and compute our detection success ratio at the level of postal code.

4. RIDE WITH THY NEIGHBOR

Here we formulate the problem of `EndPoints RS` *i.e.*, ride-sharing among people that live and work nearby. We develop a practical algorithm, apply it to the 270K users with estimated home/work locations, and compute the number of cars that can be reduced under different scenarios.

4.1 Formulation

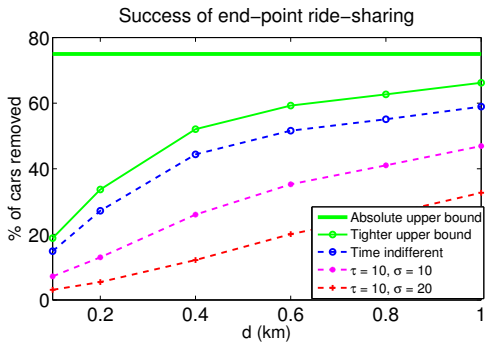


Figure 5: Benefits from `EndPoints RS`.

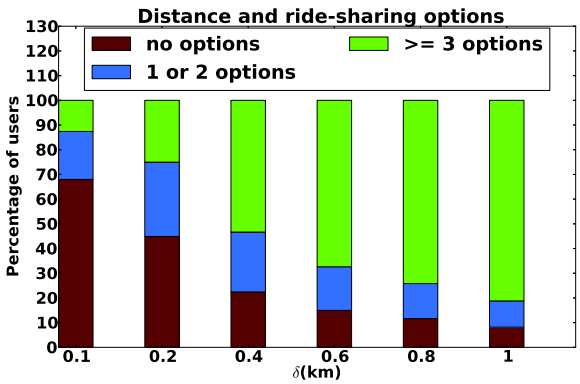


Figure 6: How δ affects the ride-sharing options

Let V denote a set of potential drivers and $c(v)$ the capacity, in terms of available seats, of the car of driver $v \in V$ and $p(v)$ a penalty paid if driver v is selected for driving his car and picking up passengers. Let $h(v,u)$ denote the geographic distance between the home locations of drivers v and u and $w(v,u)$ the corresponding distance between their work locations. Let δ denote the maximum distance between a driver's home/work and the home/work of passengers that he can pick up in his car, *i.e.*, v can have u as passenger only if: $max(h(u,v), w(u,v)) \leq \delta$

Let $d(v,u)$ denote a virtual distance between v and u defined as follows:

$$d(v,u) = \begin{cases} h(v,u) + w(v,u), \\ \text{if} \quad \max(h(v,u), w(v,u)) \leq \delta \\ \\ \infty, \quad \text{otherwise} \end{cases}$$

Our objective is to select a subset of drivers $S \subseteq V$, and find an assignment $a : V \to S$, that minimizes $P(S) + D(S)$, the sum of penalty and distance costs, while satisfying the capacity constraints of cars. The two costs are defined as follows:

$$P(S) = \sum_{v \in S} p(v) \quad \text{and} \quad D(S) = \sum_{v \in V} d(a(v), v)$$

where $a(v) \in S$ is the driver in S that is assigned to pick up passenger v (can be himself if v is selected as a driver). By setting $p(v) > 2\delta \cdot c(v)$ we can guarantee that an optimal solution will never increase the number of cars used in order to decrease the (pickup) distance cost between a driver and its passengers. The above problem is an NP-hard *Capacitated Facility Location Problem with Unsplittable Demand* in metric distance: the set of potential drivers corresponds to the set of locations; the set of chosen drivers corresponds to opened facilities; car capacity corresponds to facility capacity; distance $d(v,u)$ corresponds to the cost of assigning a location v to the facility u. Efficient approximation algorithms are known for this type of facility location problem [14].

The above formulation finds the minimum number of cars needed when there are no timing constraints around departure and return times from home and work. Next we refine the formulation to include time. We assume that departures from home and work follow Gaussian distributions, centered at 9 am and 5 pm respectively, with standard deviation σ. Also, we introduce the wait tolerance τ that captures the maximum amount of time that an individual can deviate from his normal schedule in order to share a ride, Fig. 4. More specifically, if $LH(u)$ expresses the time a person u leaves home to go to work, and $LW(u)$ expresses the time she leaves work in order to return to home. Then, two people u and v, can share a ride only if:

$$max(|LH(u) - LH(v)|, |LW(u) - LW(v)|) \leq \tau$$

The introduction of the temporal constrains will only change the virtual distance between v and u :

$$d(v,u) = \begin{cases} h(v,u) + w(v,u), \\ \text{if} \quad \max(h(v,u), w(v,u)) \leq \delta \\ \text{AND} \quad |LH(u) - LH(v)| \leq \tau \\ \text{AND} \quad |LW(u) - LW(v)| \leq \tau \\ \\ \infty, \quad \text{otherwise} \end{cases}$$

4.2 A practical algorithm

In this section we show how to modify the existing approximation algorithm [14] for the facility location problem described above and obtain a faster heuristic that can cope with the size of our data set.

The existing algorithm starts with an initial random solution and improves it iteratively via local search. At each iteration there are $O(n^2)$ candidate solutions, where n corresponds to the number of potential drivers, and for each one of them we find the assignment (passengers to drivers) that will minimize the cost; this can be done in polynomial time by solving an appropriately defined instance of the *transportation problem*. The algorithm terminates when local search cannot find a better solution.

We modify the algorithm in three ways. First, since the quality of the solution depends mostly on the number of drivers, we try to keep that number as low as possible. Therefore, we use the b-matching [3] algorithm to generate the initial solution, instead of generating it randomly. The input to the b-matching algorithm consists of the set of potential drivers V, a function $p(v)$ that defines the set options for a potential driver v *i.e.* $p(v) = \{u | d(u,v) < \inf\}$, and a global ordering of the potential drivers, O. The global ordering will be based on the number of options; the fewer the options, the higher the position in O. By using b-matching with a global order we are guaranteed to find a solution in $O(n)$ time [3]. For each match generated by b-matching, we assign the potential driver with the most occupied seats to drive; we make sure that every user in V appears in only one car. This solution proves much better than the random

one by paying $O(nlog(n))$ for sorting the users to generate the global preference list and $O(n)$ for the matching.

Second, solving a transportation problem with 270K users is hard. Therefore, we need to modify the local search steps of the approximate algorithm. Given an initial solution we leave the users commuting in cars of four as they are and search for better assignments only for the rest. This way the size of the transportation problem will be reduced and that would speed up the process of generating the assignment.

Third, reducing the size of the transportation problem is not enough; we also need to reduce the neighborhood of candidate solutions. Given an initial set of drivers, S, we create a fixed size neighborhood, where each solution S' is created by doing random changes in S. The reason why we do that is because considering all potential solutions that differ from S only by one, means that we have to examine $O(n^2)$ candidate solutions; that makes each iteration very expensive. Therefore, the fixed size solution helps us speed up the time we spend in each improvement step.

Without the above modifications it would be impossible to solve the problem in real time. Solving an instance of the transportation problem for 270K users required a couple of hours for $\delta = 0.6$ km, and even more when $\delta = 0.8$ or $\delta = 1.0$ km. Therefore, solving $O(n^2)$ such problems for a single iteration becomes too expensive. Moreover, most of the time the solution of the b-matching algorithm was so good that the gain from the improvement steps would be insignificant.

4.3 Results

A this point we are ready to calculate the effectiveness of `EndPoints RS` in the Madrid metropolitan area. We reduce the size of our dataset by randomly selecting only 60% of the users.We do that to capture the fact that only 60% of the population has a car in the area of Madrid [1]. We also show results for the case that half of the car owners use their car at their daily commute (the results are quantitatively close). For the remaining of the section, we will refer to users who can share rides with a specific user v, as *options of* v. Subsequently we compute the percentage-wise reduction of cars

$$\text{success} = \frac{\#(\text{init. cars}) - \#(\text{ride-sharing cars})}{\#(\text{init. cars})} \cdot 100$$

using the following algorithms:

Absolute upper bound: Given our definition of success, we cannot do better than 75%. This is the case when all cars carry 4 people.

Tighter upper bound: All users with at least one ride-sharing option, are assumed to commute in cars of 4.

Time-indifferent matching ($\tau = \infty$): This is the practical algorithm described in Sect. 4.2

Time-aware matching: This is the version of the algorithm that considers timing constraints under the assumption of normally distributed departure times.

In Fig. 5 we see what happens when the drivers are willing to tolerate a detour of δ and deviate τ minutes from their departure times, in order to share the same car with another individual. The results show that with even modest delay tolerance of 10 minutes and detour distance of 0.6 Km (a couple of city blocks) 40% of the cars can be saved. This is more than half of the absolutely optimal performance. Increasing either of the two parameters improves the success ratio. The diminishing improvement with increasing δ can be explained by the number of options users have given the distance δ. In Fig. 6 the red color represents the users with no options, the blue color the users with 1 or 2 options, and the green color the users with 3 or more options. We can see that the success of ride-sharing is proportional to the number of users who have 3 or more options; since, as we can see from Fig . 5 and Fig . 6, they increase in a similar way.

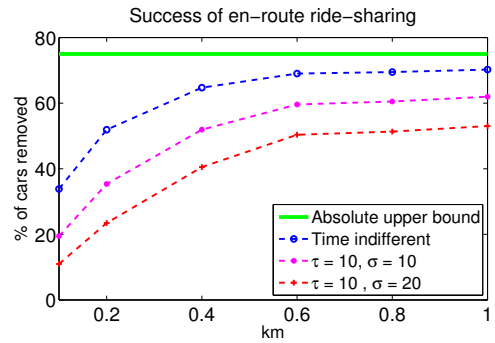


Figure 7: Benefits from `EnRoute RS`.

5. EN-ROUTE RIDE-SHARING

The effectiveness of ride-sharing can be greatly enhanced by picking up additional passengers en-route. For example a driver that lives in a sparsely populated area might not have any neighbors to fill his seats but once he enters the city he might be able to pick several passengers that have routes "covered" by his own. To quantify the benefits of en-route ride-sharing we obtain routes from Google Maps for our dataset of 270K users and extend the algorithm of Section 4.2.

5.1 En-route algorithm

We use an iterative algorithm with the following steps in each iteration.

1. Run the basic `EndPoints RS` algorithm.
2. Exclude from the solution cars that get fully packed (a car of 4). Then order cars in decreasing order of passengers and start "routing" them across Madrid using data from Google maps.
3. When the currently routed car v meets a yet un-routed car v', then v is allowed to steal passengers from v' as long as it has more passengers than v' (a rich-get-richer strategy). Whenever a routed car gets fully packed it is removed from further consideration. Whenever a car with a single passenger is encountered the number of cars is reduced by one.

This is repeated until there is no possible improvement. It can be shown (omitted for lack of space) that the rich-get-richer rule leads to convergence.

5.2 Results

Fig. 7 shows the performance of `EnRoute RS`. To make the comparison with `EndPoints RS` easier we summarize results from both in Table 1: one can verify the significant improvement obtained through `EnRoute RS`, which is several cases comes within 10% of the optimal performance.

6. RELATED WORK

Mining mobility patterns from wireless traces has recently received a lot of attention. Of particular interest to this work is the mining of mobility patterns from CDRs. The best examples of this area are the work of Gonzalez et al. [8], who use CDRs in order to characterize the distribution of human trajectories, and the work of Isaacman et. al. [10] [11] [12] [13] who use CDRs to characterize various aspects of human mobility, such as important places, ranges of travel, etc. The previous examples use only the location information, but recent work [4] [6] also exploits the social graph inside the CDRs also (who calls who). Finally, another example of human mobility form CDR data is [7].

To the best of our knowledge this is the first work on ride-sharing applications based on CDRs. Other ride-sharing applications, that exploit wireless traces, use mostly GPS [9] [19]. The work presented in [9] presents a frequency-based route mining algorithm designed for GPS data and is used to extract frequent routes and recommend ride-sharing plans using the GPS traces of 178 individuals. Trasarti et al. [19] use GPS data to build the mobility profiles of 2107 individuals and match them based on the similarities of their mobility profiles; they also apply their algorithms to a GSM-like data set, which they synthesize by reducing the size of their GPS data set.

Past work on Carpooling has be focused mainly in characterizing the behavior of carpoolers, identifying the individuals who are more likely to carpool and explaining what are the main factors that affect their decision on whether to use carpooling or not [16]. Also, the question whether ride-sharing can reduce congestion has been asked before [20]. But, the authors of [20] assumed uniform distribution of home and work locations and concluded that ride-sharing has little potential for congestion reduction. On the contrary, we make no assumption about the distribution of home and work locations in a city, but infer such information from a Call Description Dataset, and show that ride-sharing has a lot of potential for reducing the number of cars from the streets of a city.

7. CONCLUSION

In this paper, we used CDR data to derive an upper bound for the potential of ride-sharing. The results indicate that there exists huge potential in densely populated urban areas, such as the city of Madrid in Spain. This motivates working on the technological challenges involved in facilitating car sharing. In future work, we plan to (i) extend our study in other cities, (ii) take additional aspects into account, such as the structure of the call graph, (iii) obtain additional information for ground truth and (iv) deploy an actual system that mines CDR and facilitates car sharing.

Table 1: Summary of results, $\delta = 0.6$ km

drivers (%)	δ (km)	τ (min)	σ (min)	`EndPoints RS` (%)	`EnRoute RS` (%)
100	0.6	–	–	**56**	**69**
100	0.6	10	10	**41**	**60**
100	0.6	10	20	**26**	**50**
60	0.6	–	–	**52**	**67**
60	0.6	10	10	**35**	**56**
60	0.6	10	20	**20**	**46**
30	0.6	–	–	**44**	**63**
30	0.6	10	10	**27**	**50**
30	0.6	10	20	**13**	**40**

The first column indicate the % of users owning a car, the second indicates the distance tolerance, the third parameter indicates the time tolerance, the fourth the time spread and the last two columns show the success of carpooling for `EndPoints RS` and `EnRoute RS`.

8. REFERENCES

[1] Instituto de estadistica de la comunidad de madrid: http://www.madrid.org/iestadis/.

[2] The true cost of a car over its lifetime: http://www.doughroller.net/smart-spending/true-cost-of-a-car-over-its-lifetime.

[3] K. Cechlárová and T. Fleiner. On a generalization of the stable roommates problem. *ACM Trans. Algorithms*, 1(1):143–156, July 2005.

[4] Eunjoon Cho, S.A. Myers, and J. Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proc. ACM SIGKDD*. ACM, 2011.

[5] Nigel Davies, Manfred Lau, Chris Speed, Tom Cherrett, Janet Dickinson, and Sarah Norgate. Sixth sense transport: challenges in supporting flexible time travel. In *Proceedings of the Twelfth Workshop on Mobile Computing Systems and Applications*, HotMobile '12, pages 8:1–8:6. ACM, 2012.

[6] N. Eagle, A. Pentland, and D. Lazer. Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences of the United States of America*, 106(36):15274–8, September 2009.

[7] M. Ficek and L. Kencl. Inter-Call Mobility Model: A Spatio-temporal Refinement of Call Data Records Using a Gaussian Mixture Model. In *Proc. Infcocom*, 2012.

[8] M. C. González, C. A. Hidalgo, and A. L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–82, June 2008.

[9] W. He, D. Li, T. Zhang, L. An, M. Guo, and G. Chen. Mining regular routes from gps data for ridesharing recommendations. In *UrbComp*. ACM, 2012.

[10] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. Identifying Important Places in People's Lives from Cellular Network Data. In *Proc. Pervasive*, June 2011.

[11] S. Isaacman, R. Becker, R. Caceres, S. Kobourov, M. Martonosi, J. Rowland, and A. Varshavsky. Ranges of human mobility in Los Angeles and New York. In *Proc. MUCS*, March 2011.

[12] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, J. Rowland, and A. Varshavsky. A tale of two cities. In *Proc. HotMobile*, February 2010.

[13] S. Isaacman, R. Becker, R. Cáceres, M. Martonosi, J. Rowland, A. Varshavsky, and W. Willinger. Human Mobility Modeling at Metropolitan Scales. In *Proc. MobiSys*, June 2012.

[14] Madhukar R. Korupolu, C. Greg Plaxton, and Rajmohan Rajaraman. Analysis of a local search heuristic for facility location problems. In *Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 1998.

[15] V. Soto and E. Frias-Martinez. Automated Land Use Identification using Cell-Phone Records. In *HotPlanet*, 2011.

[16] R.F. Teal. Carpooling: Who, how and why. *Transportation Research*, 21A(3):203–214, 1987.

[17] Arvind Thiagarajan, Lenin S. Ravindranath, Hari Balakrishnan, Samuel Madden, and Lewis Girod. Accurate, low-energy trajectory mapping for mobile devices. In *8th USENIX Symp. on Networked Systems Design and Implementation (NSDI)*, Boston, MA, March 2011.

[18] Arvind Thiagarajan, Lenin S. Ravindranath, Katrina LaCurts, Sivan Toledo, Jakob Eriksson, Samuel Madden, and Hari Balakrishnan. Vtrack: Accurate, energy-aware traffic delay estimation using mobile phones. In *7th ACM Conference on Embedded Networked Sensor Systems (SenSys)*, Berkeley, CA, November 2009.

[19] R. Trasarti, F. Pinelli, M. Nanni, and F. Giannotti. Mining mobility user profiles for car pooling. In *Proc. UrbComp*. ACM, 2011.

[20] H.-S. J. Tsao and D. Lin. Spatial and temporal factors in estimating the potential of ride-sharing for demand reduction. California PATH Research Report, UCBITS-PRR-99-2, 1999.