Taxi Exp: A Novel Framework for City-Wide Package Express Shipping via Taxi Crowd Sourcing

Conference Paper · December 2014

DOI: 10.1109/UIC-ATC-ScalCom.2014.112

Chao Chen
Chongqing University

Daqing Zhang
Institut Mines-Télécom

Leye Wang
The Hong Kong University of Science and Technology

Xiaojuan Ma
Huawei Technologies

TaxiExp: A Novel Framework for City-wide Package Express Shipping via Taxi CrowdSourcing

Chao Chen[†], Daqing Zhang[‡], Leye Wang[‡], Xiaojuan Ma[♭], Xiao Han[‡], Edwin Sha[†]
[†] College of Computer Science, Chongqing University, Chongqing 400044, China
[‡] CNRS SAMOVAR, Institut Mines-TELECOM/TELECOM SudParis, Evry 91011, France
♭ Huawei Noah's Ark Lab, Hong Kong, China.

***Abstract*—Despite the great demand on and attempts at package express shipping services such as the same-day delivery feasible for online firms, turning a profit is still difficult. To develop more economical or even cost-free transportation of packages, in this paper, we propose to make use of the existing taxis on the street that are delivering passengers, in a crowd-sourced manner. To the best of our knowledge, this is the first work that exploits taxis occupied by passengers to help deliver package collectively, without hurting the quality of taxi services. Specifically, we propose a two-phase framework for the package express shipping. In the first phase, we rank the road segments according to their *influential factor* values, which is similar to the idea of identifying key people in social networks. Hubs are then identified based on the ranking and the geographical locations of the road segments. In the second phase, we develop two inter-hub routing algorithms, namely, *First-Come-First-Service (FCFS)* and *Destination-Closer (DesCloser)*, to ship a package to its destination. We evaluate the two-phase framework on a large-scale real-world taxi data set, generated by 7,600 taxis in a month. Results show that, on average, the package delivery time based on *DesCloser* is 5.3 hours, which is 2.6x shorter than that of *FCFS*; the number of participating taxis per package based on *DesCloser* is 3.10, which is 10.6x fewer than that of *FCFS*.**

Keywords*—*Package Shipping; Online Business; Opportunities; Taxi CrowdSourcing

I. Introduction

Nowadays online shopping via Internet has become very popular, probably due to the attractive price and great convenience. According to a report by Forrester Research Incorporated[1], online retail sales will reach $327 billion in the United States in 2016, a significant increase from the $226 billion in 2012 [16]. However, compared to traditional retailing, online retailing still suffers from several disadvantages. One key weakness is that customers cannot get their products *instantly*. People have to be patient while their packages are on the way, and often get frustrated due to unexpected delay in delivery. The delivery time is a major concern for online consumers when they have immediate needs. Therefore, more and more online firms including giants (e.g. eBay[2], Google[3], Amazon[4], Walmart[5]) and startups (e.g. the Fancy[6]) are experimenting with the *same-day delivery* service [20] to make their retail business more competitive in the market.

TABLE I. Pricing of same-day delivery from different online firms.

Online Firm	Service offered	Pricing*
eBay	eBay now	$5 additional fee per order
Google	Google Shopping Express	$5 additional fee per order
Amazon	amazon fresh	$5.99 additional fee per order
Walmart	Walmart to Go	$8.99 additional fee per order

*The price here may not be updated and it also varies at different cities. Note that some companies have additional constraints on orders, or have special discounts to membership. For example, eBay only provides such service on orders of more than $25; Amazon offers free shipping on orders of more than $35 to memberships.

Costs come with speeding up the shipping process, and thus business providers often charge customers an additional fee if they want their products to be available in a tight deadline. The price per order is not cheap as shown in Table I, which intimidates many customers. What is worse, such service is limited to a few big cities only. It is because, to save time, online firms often have to work closely with local retailers/shops to dispatch products *locally*. In addition, online firms have been looking into more economical means of transportation to reduce the cost of the delivery.

Despite the needs, doubts regarding whether same-day delivery can succeed in the long run arise constantly [14], as there has not yet been any effective solution to cutting the operational cost and increasing the profit margins. *High transportation cost is one of the main bottlenecks.*

To develop more economical or even cost-free shipping channels, in this paper, we propose to have packages take *hitchhiking rides* with existing taxis that are delivering passengers on the street. We illustrate the basic idea by the following use case, as shown in Fig. 1. There is a package to be delivered from **A** to **B**. Opportunistically, there happens to be a passenger (i.e. *Passenger 1*) near **A** (i.e. at **C**) making a real-time taxi ordering request[7], intending to go to **D**. Once a taxi responds to the request (*Taxi 1* in this case), we can assign the package delivery task to its driver. More specifically, on *Taxi 1*'s way to pick up *Passenger 1* at **C**, we can first ask the driver to collect the package at **A**. Then, the package will be sent to **D**, together with *Passenger 1*. After dropping off *Passenger 1*, the taxi driver (*Taxi 1*) is required to leave the package in the safe-boxes in **D**. The package has not yet arrived at its destination in this case with the

[1] http://www.forrester.com/home/
[2] http://www.ebay.com
[3] http://www.google.com
[4] http://www.amazon.com
[5] http://www.walmart.com
[6] http://fancy.com

[7] It is popular that passengers order taxis in real time with mobile apps such as Uber (https://www.uber.com/). Usually, to make a request, a passenger has to provide information including his/her origin, intended destination. The request will be broadcast locally, and the taxi driver who responses the request would come to pick up the passenger.

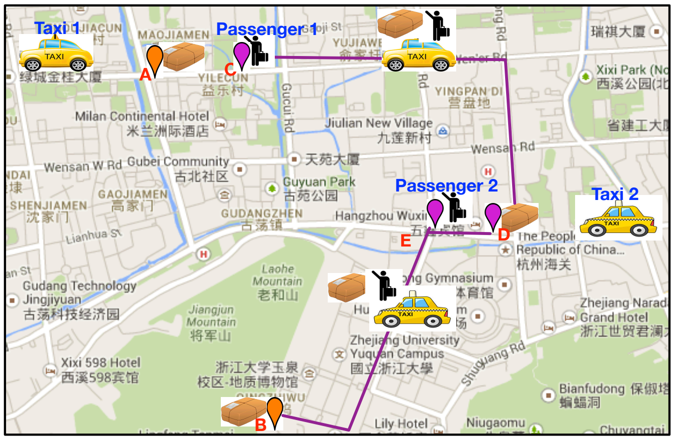

Fig. 1. A package is delivered to its destination via taxi crowdsouring. Two taxis are recruited to participate the package delivery task in the example.

first hand-off of *Taxi 1*, and thus we continue to search for the next ride in a similar manner. In this example, after the second hand-off of *Taxi 2*, the package gets to its destination, together with *Passenger 2* who departs at **E**. As can be seen from the use case, the only extra costs occur between the pick-ups of package and then passenger as well as between the drop-offs of passenger and then package. Compared to conventional dedicated shipping channels, delivering packages via taxi crowdsouring has several advantages:

Abundant opportunities. Currently, taxi has become one of the main transportation modes for commute and travel. Taking Hong Kong as an example, there are 15,000 taxis by the end of 2013 and they serve more than 1,000,000 passengers everyday [1], providing many *opportunities* to help hand-off packages in the city.

Short delivery time. As suggested in [11], [12], [22], [24], taxi drivers have excellent knowledge of the traffic conditions at different roads at different time, and they are often good at choosing paths with lighter traffic to deliver passengers to their destinations. Hence, it is highly possible that the time cost of packages on roads is short as we rely on the taxis when they are delivering passengers.

Economical and environmental friendly. Compared to introducing additional transportations that are dedicated to package delivery, we leverage the existing taxis on the street. Moreover, we only assign packages to taxis that are carrying passengers to reduce extra air and noise pollution, making it an economical, green solution.

To make taxi crowdsourcing feasible, we need to address the following research challenges.

1) **How to ensure the quality of taxi service?** Participating in crowdsourced package express shipping should not impair the quality of passenger delivery service, which is the core business of the taxi industry.

2) **Where to store the package in each hand-off?** Often, a package cannot be arrived at its destination on a single taxi ride. It has to be dropped off at certain "exchange station," e.g. despatch center, warehouse, or simply safe boxes, and wait for the next ride. The numbers and locations of such hand-off points may impact the efficiency of the entire express shipping service.

3) **How to minimize the total package delivery cost?** Package delivery time not only depends on the total travel time on the roads, but also the waiting time at the safe-boxes between hand-offs. It is non-trivial to model the waiting time, as it varies with storing locations and time of the day, and also depends on the destination of the package.

In this paper, we make the following contributions:

1) To the best of our knowledge, this is the first attempt to achieve economical and environmental friendly package express shipping via taxi crowdsourcing. To better allocate delivery tasks to existing taxis while ensuring the quality of their passenger service, it is better to know where taxis will pick up passengers and where they are heading, which is proven to be quite difficult by previous work [21]. To simplify this problem, we only involve taxis which take orders from real-time taxi requesting apps such as Uber, DiDi and KuaiDi, where such information is included.

2) We propose a two-phase framework to solve the problem. The first phase is *hub identification*, and the second phase is *inter-hub routing*. Specifically, in the first phase, we adopt the idea of identifying influential people in social networks to rank the influence of road segments in the road network, based on taxi trajectories *in history*. Then we identify hubs based on the ranking and the locations of road segments. In the second phase, we propose two inter-hub routing algorithms (i.e. *First-Come-First-Service* and *Destination-Closer*) to ship packages between any Origin-Destination (OD) pair. We schedule taxis that respond to taxi ordering requests, to participate the delivery tasks *in real time*.

3) We evaluate the proposed framework extensively using large-scale real-world taxi data sets. Results show that, even with simple routing algorithms, the two-phase framework can achieve good performance. Specifically, over 98% of packages can arrive at their destinations by the deadline (i.e. before the end of the day). On average, the package delivery time based on *DesCloser* is 5.3 hours, which is 2.6x shorter than that of *FCFS*; the number of participating taxis per package based on *DesCloser* is 3.10, which is 10.6x fewer than that of *FCFS*.

The rest of the paper is organized as follows. In Section II, we introduce some key definitions and the assumptions we have made in this paper. In Section III, we present the details about our two-phase framework. We evaluate the performance of the proposed framework in Section IV. In Section V, we review the related work and show how this paper differs from prior research. Finally, we conclude the paper and propose the future directions in Section VI.

II. Definitions and Assumptions

In this section, we present some key definitions, followed by of the assumptions we have made.

A. Definitions

Definition 1 (Road Network) A road network is a graph, where nodes are the endpoints of road segments, and edges are road segments. Each node in the road network is represented by its latitude/longitude values. For simplicity, each edge here is represented by the average latitude/longitude values of its two endpoint nodes.

Definition 2 (Neighbouring Road Segments) For a given road segment, its neighbouring road segments are defined as the set of road segments that share the same node with itself in the road network, which is denoted as $Neighbour(\cdot)$.

Definition 3 (Package Delivery Query) A package delivery query is defined as a triplet $\langle O_p, D_p, T_p\rangle$, where O_p and D_p refer to the origin and destination of the package respectively; T_p refers to the time when the user (e.g. the online buyer) submits the request (i.e. the birth time).

Definition 4 (Taxi Ordering Query) A taxi ordering query is defined as a triplet $\langle O_t, D_t, T_t\rangle$, where O_t and D_t refer to the passenger's origin and his/her intended destination. T_t refers to the time that the passenger submits the request.

Definition 5 (Link Graph) A link graph $G(V, E)$ is a directed graph built based on the passenger-delivery trajectories in history, where V is a set of road segments, E is a set edges. The nodes defined in the link graph are actually the edges in the road network. Specifically, each edge e_{ij} refers to the number of taxi trajectories from the road segment i to road segment j. Note that we map the pick-up/drop-off points to the road network, and represent each taxi trajectory as a pair of road Ids where the passenger was picked up and dropped off.

Definition 6 (Graph and Node Entropy) Graph and edge entropy are computed on the link graph $G(V, E)$. Eq. 1 defines the graph entropy, and Eq. 2 defines the node entropy respectively.

$$EnG(G) = -\sum_{i=1}^{|V|} p(v_i)log(p(v_i)) \tag{1}$$

where $p(v_i) = \frac{\sum_{j=1}^{|V|} e_{ij}}{\sum_{i=1;j=1}^{|V|} e_{ij}}$.

$$EnNode(v_i) = -\sum_{j=1, j\neq i}^{|V|} p(e_{ij})log(p(e_{ij})) \tag{2}$$

where $p(e_{ij}) = \frac{e_{ij}}{\sum_{j=1}^{|V|} e_{ij}}$.

B. Assumptions

We make the following assumptions in the paper.

Assumption 1: Taxi drivers are willing to accept the assigned package delivery tasks.

We believe this assumption is realistic since we can design prober incentive mechanisms. In the design of incentive mechanisms, a prime principle is to ensure more tasks that a taxi participates, more money should be rewarded. However, this problem is beyond the scope of this paper.

Assumption 2: Most of road segments are equipped with safe-boxes, which can be used to store packages *temporally*.

This assumption also makes sense. We can install safe-boxes at the taxi stands, which does not incur too much extra cost in terms of installing and maintenance.

Assumption 3: The package can be *trackable*.

Since the birth time of the package, it is either stored at the safe-box or transported by an assigned taxi. Each safe box has a unique Id; each taxi driver is registered in taxi management department and also has a unique Id. So during the delivery process, this package can always be tracked back to an ID (either safe-box or taxi driver), i.e., being trackable.

Assumption 4: A one-time security code is required to open the safe-box. The code will be sent to the mobile phone of the scheduled taxi driver, or sent to the mobile phone of the receiver for pick-up when the package arrives at its destination.

The last two assumptions try to address the package security issues, since besides being fast and affordable, security is another key concerns of a shipping service.

III. The Two-Phase Framework

In this section, we will discuss the details about our proposed two-phase framework, i.e. *hub identification* and *inter-hub routing*. When we take public transportation, if a destination is not directly reachable, we have to go to some interchange stations first. Sometimes, we have to make several stops. We can apply similar mechanism to package delivery. To this end, in the first phase, we aim to identify hubs (similar to the interchange stations) based on the taxi trajectories in history. This phase is conducted *offline*. The second phase is to develop online inter-hub routing algorithms based on the real-time taxi ordering and package delivery queries.

A. Hub Identification

There are three important requirements that desirable hubs should meet. First of all, a hub should be popular among taxi and passengers (i.e. with dense pick-ups and drop-offs). Second, from the hub, many places can be reached by taxi (i.e. having large in-coming and out-going degrees). Finally, hubs around the whole city should be distributed properly to speed up the efficiency (i.e. the total package delivery time). Taking the public transportation scenario as an example again, interchange stations usually locate in different corners in the city (e.g. center, north, south), and are far-away from each other.

We introduce an entropy-based algorithm to rank different road segments according to their *influential factor* (i.e. both popularity and accessibility), similar to locating key players in a social network [19]. The pseudo-code is shown in Algorithm 1. First of all, we identify top-k road segments with the top-k number of sum of pick-ups and drop-offs. Nodes which do not belong to the top-k group will be removed, and their connecting edges will be deleted as well. In other words, only top-k road segments are selected to build a new link graph (Lines 1∼4). Lines 4∼10 show the process of calculating the *influential factor* value for each road segment. Finally, the algorithm generates a list of road segments ranked by the *influential factor* values.

However, it is possible that some high-influential road segments would locate closely together and concentrated in

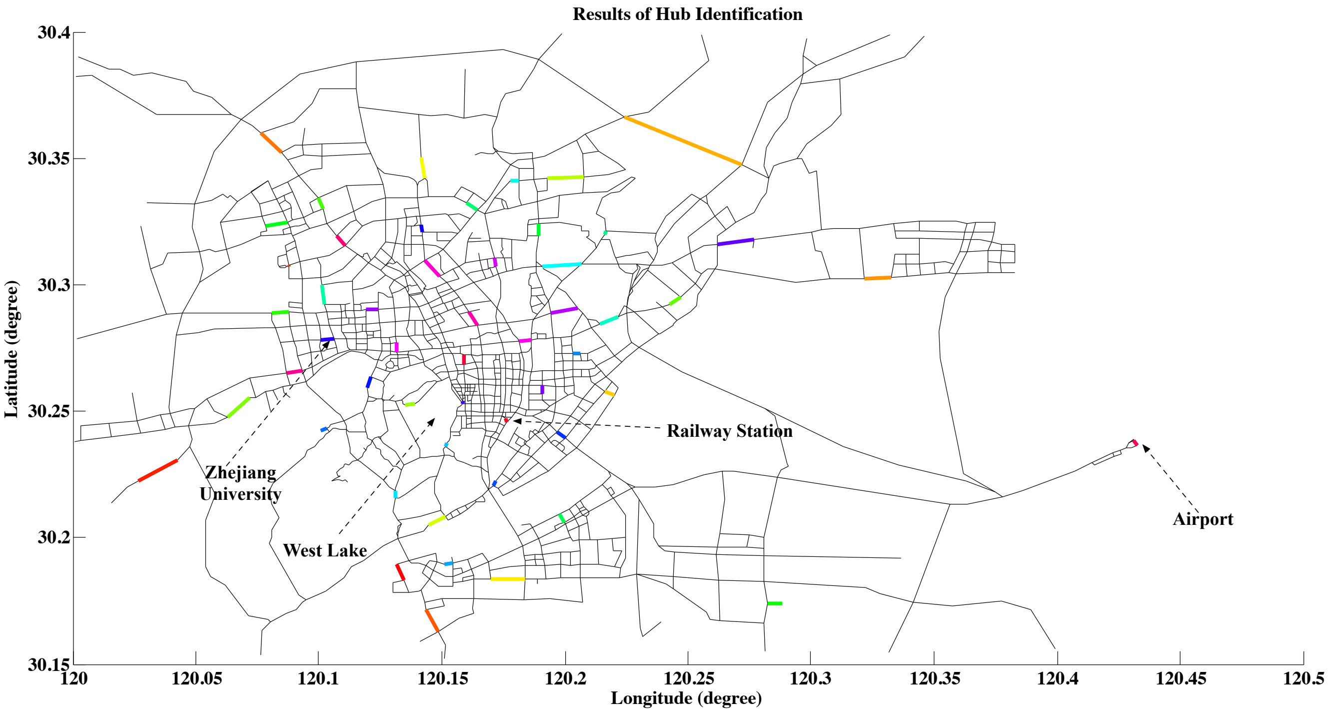

Fig. 2. Hub identification results. The road segments marked in different colors are the identified hubs. (Best viewed in the enlarged digital version)

Algorithm 1 Road Segments Ranking Algorithm.

Input: Graph $G(V, E)$ built by historical taxi trajectories (with n nodes in total);
Output: A set of k nodes.

for $i := 1$ **to** n **do**
 $freq(i) = \sum_{j=1}^{n} e_{ij}$; //Calculate the total pick-ups and drop-off of node v_i'.
end for
Rank nodes according to the $freq$ values, and select the top-k nodes to build the new graph $G(V', E')$.
for $i := 1$ **to** k **do**
 Calculate the edge entropy of node v_i' according to Eq. 2.
 Drop node v_i' and its connected edges from the graph.
 Calculate the graph entropy of the remnant graph (G') according to Eq. 1.
 $IF(v_i') = \frac{EnNode(v_i')}{EnG(G')}$ //Calculate the *Influential Factor* (IF) of the given node v_i'.
end for
Rank nodes according to IF values.

Algorithm 2 Hub Identification Algorithm

Input: A list of road segments ranked by the *influential factor* (V).
Output: A set of hubs.

1: $i = 1$;
2: $hubs = \emptyset$
3: **while** $V \neq \emptyset$ **do**
4: $hubs = \{v_i\} \cup hubs$;
5: $V = V \backslash \{v_i\}$ // Remove V_1 from V
6: $k = |V|$;
7: **for** $j := 1$ **to** k **do**
8: **if** $dist(v_i, v_j) < Th$ **then**
9: $V = V \backslash \{v_j\}$ //Remove v_j from V
10: **end if**
11: **end for**
12: $i = i + 1$;
13: **end while**

small down-town area, since the ranking algorithm does not consider the relative locations of nodes, which fails to meet the last requirement. To better allocate the hub resources around the whole city, we further take the geographic locations of road segments into consideration, as shown in Algorithm 2. The basic idea is similar to our previous work on bus stop identification [6]. We *iteratively* select the most influential road segment as a hub and remove road segments close to it (Line 4∼11). As a result, we keep only the most influential road segment in each given region (determined by the threshold in Line 8).

Results shown in Fig. 2 are the hubs identified by the proposed algorithms. Note that the results are obtained based on one-month taxi GPS data in the city of Hangzhou, China. In total, we have identified 52 hubs in the whole city. To better understand the distribution of hub locations, we also marked some landmarks and the down-town area (near the West Lake) in the city of Hangzhou. As expected, we can see that identified hubs almost distributed evenly around the whole city.

B. Inter-hub Routing

We propose two algorithms for inter-hub routing. The first algorithm is the *First Come First Service (FCFS)*. The basic idea of the *FCFS* is: as the package delivery and taxi ordering queries ($\langle O_p, D_p, T_p \rangle$ and $\langle O_t, D_t, T_t \rangle$) arrive in stream, for each package delivery query, we assign the package to the first taxi[8] which will pick up a passenger at the same road segment (or its neighbouring road segments) after the birth time of the package, and will head to one of the hubs (or the neighbouring road segments of hubs), *regardless of the destination of the package*. Mathematically, it should satisfy: $O_t \in Neighbour(O_p); T_t > T_p; Neighbour(D_t) \in \{hubs\}$. After the first hand-off, the package will be stored at one of the hubs until the next suitable taxi collects it($O_p \leftarrow hub_i$), following the *FCFS* algorithm. Inter-hub routing terminates when the package is offloaded to a hub close to its destination (e.g. with distance less than a predefined threshold). Finally, we will wait to assign the package delivery task to the taxi which will go to the destination of package directly. It should be noted that during the inter-hub routing, we only assign the delivery tasks to taxis the origin and destination of which are both a hub or near a hub. The only exception is that if the taxi would go to the destination (i.e., the destination of its picked-up passengers) exactly the same as the destination of the package. In this case, the package will be assigned to that taxi directly, and no more inter-hub routing is necessary.

The second algorithm is *Destination-Closer (DesCloser)*. The basic idea of *DesCloser* is to assign the package to the first near-by taxi which will head to the hubs closer to the destination of the package, *compared to the distance from current location of package to its destination*. In other words, no matter where the package is originated, after each hand-off, it will reach one of hubs, closer to its destination. Similarly, taxis from the current location of the package but not heading to anywhere close to any hubs might still be assigned if and only if they are going to the destination of the package. In most cases, inter-hub routing will terminates when the package arrives at a hub with a distance to the destination shorter than a predefined threshold. Similar to the *FCFS* algorithm, after the package stored at the last hub, we will wait and assign the package delivery task to the taxi which will start from that hub to the destination of package directly.

IV. Experimental Evaluation

A. Experiment Setup

Data Description. We validate the proposed two-phase framework with a large-scale real-world taxi GPS dataset, generated by 7,600 taxis in a month in a large city in China (Hangzhou). We only keep trajectories of taxis occupied by passengers.

Query Simulation. As discussed in Section III, the first phase (i.e. *hub identification*) computed offline using the historical taxi trajectory data; However, the second phase (i.e. *inter-hub routing*) requires the real-time package delivery and taxi ordering queries as input, which was not available at the time of the evaluation. Instead, we simulate both of the real-time package and taxi ordering queries, using the following method.

For the *real-time package delivery queries*, we randomly generate their origins, destinations, and birth time. In total, we

[8] For brevity, in the rest of presentation, without special explanation, when we mention the "taxi", it refers to the taxi which will pick up a passenger at the road segment where the package located or its neighbouring road segments.

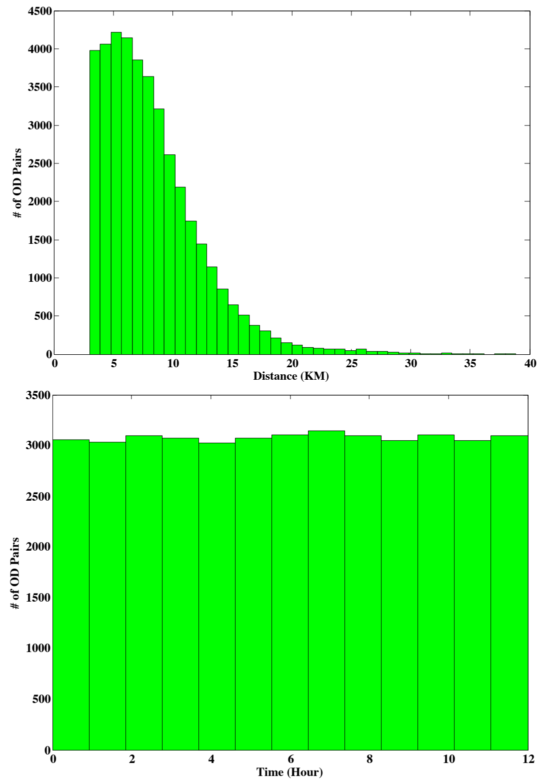

Fig. 3. Histogram for the distance (top) and time (bottom) of 40,000 OD pairs of the package.

generate 40,000 OD pairs. Fig. 3 shows the statistical information about the generated OD pairs, including the distance and the birth time. Most of them are with distance in the range of [3 10] kilometers, and the birth time almost distributes evenly in the range of [0 12] o'clock. In the following evaluations, we equally divide 40,000 in to 8 groups, each with 5,000 OD pairs.

For the *real-time taxi ordering queries*, we can simply sample and "replay" some of taxi trajectories in history since they all include the time-stamped GPS information. In other words, we use the pick-up, drop-off road segments and pick-up time of the selected taxi trajectories as the taxi ordering queries.

Evaluation Environment. All the evaluations in the paper are run in Matlab on an Intel Xeon W3500 PC with 12-GB RAM and Windows 7 operation system.

B. Evaluation Metrics

We measure the performance using the following metrics:

Success Rate: the percentage of packages that *successfully* arrive at their destinations before the deadline. Note that the *deadline* is set to be the end of a day (i.e. 23:59:59) for all the experiments, regardless of the submission time of the package requests.

Average Package Delivery Time: the average time per package between receiving the request and *successfully* reaching its destination.

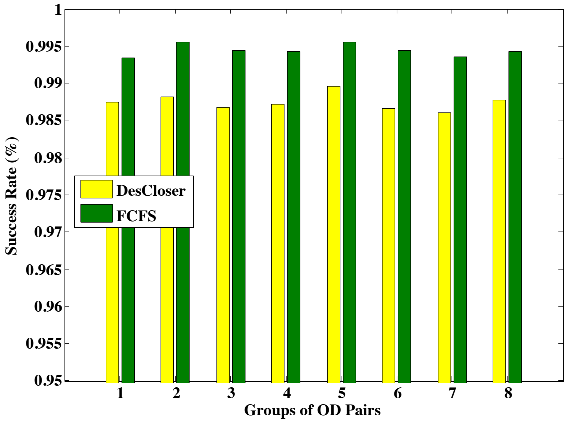

Fig. 4. Comparison result of success rate for two different inter-routing algorithms.

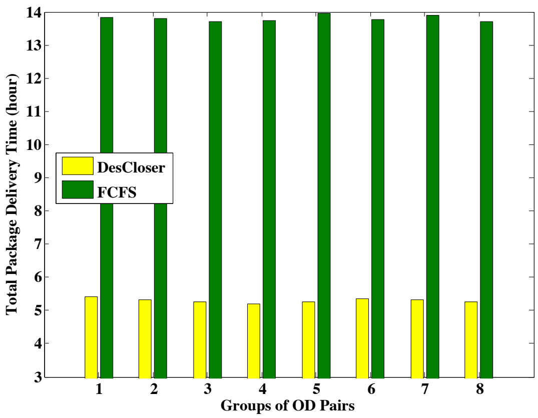

Fig. 5. Comparison result of the average delivery time for two different inter-routing algorithms.

Average Number of Participating Taxis: the number of participating taxis per package that is *successfully* delivered.

C. Performance Analysis

Fig. 4 shows the comparison of the *success rate* metric between the two inter-routing algorithms. Out of all eight groups of OD pairs, both of algorithms achieve success rate greater than 98%. The success rate of *FCFS* is even higher, with all values greater than 99%. The success rate of *DesCloser* is lower than those of *FCFS* in each group of OD pairs. This is probably because: with the *DesCloser* algorithm, the inter-routing may stopped at the hub very close to the destination of the package, however, no passenger would order the taxi from that hub to the destination of the package due to the short distance. For this case, we may ask a free taxi to help send the package to its destination in the future if the failure is not acceptable. We further perform a statistical study about the road segments where the package starts or ends are often failed (cannot be delivered successfully). As expected, these road segments are far-away from city centers, and are seldom visited by taxis.

Fig. 5 shows the comparison of the *average package delivery time* between the two inter-routing algorithms. We can see that the average delivery time of *DesCloser* is much shorter than that of *FCFS* for all 8 groups. On average, with the *DesCloser* algorithm, the total time spending on the package delivery is around 5.3 hours, compared to around 13.75 hours

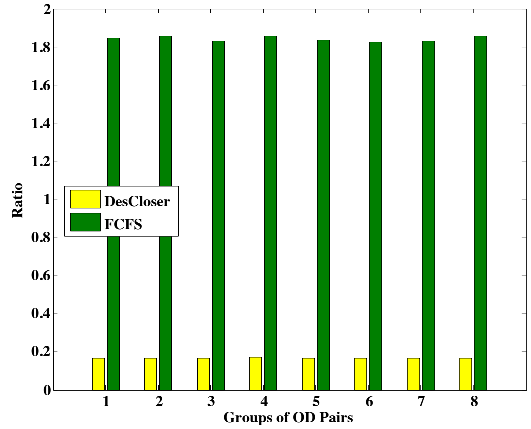

Fig. 6. Comparison result of the average ratio for two different inter-routing algorithms.

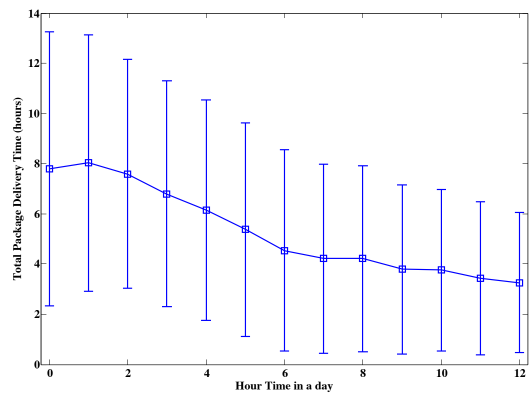

Fig. 7. Time distribution on the hour time of a day.

for *FCFS*, which is a 2.6x difference. It is because with the *FCFS* algorithm, the package may move *back and forth* during routing; while with the *DesCloser* algorithm, the package always moves *towards its destination* gradually. We further show the distributions of the package delivery time given the different birth time of the package (the time of the day) of the *DesCloser* strategy in Fig. 7, with standard deviations. The mean package delivery time almost decreases with the birth time of the package, which is reasonable since few people take taxis in the early morning and taxis that can be recruited are quite limited. The big variation in delivery time is probably due to the varied sources and destinations of the packages. OD pairs which are frequently visited by taxis are more likely to take less time; while some OD pairs just cannot be covered by the deadline, as discussed in the previous paragraph.

The total package delivery time is the sum of the time on the roads and the time waiting in the safe-boxes (i.e. storing time). We are thus interested in the *ratio* of the time spent on the roads to that in the safe-boxes, as shown in Fig. 6. We conclude that, for the *DesCloser* algorithm, packages sit in the safe-boxes most of the time (the average road-box ratio is only 0.16), since we have to wait for the *suitable* taxis that will go to the hubs closer to the destination of the packages, which is relatively infrequent. While for the *FCFS* algorithm, the moving time is almost 1.84x of the sitting time, since any taxis heading to any hubs can be recruited.

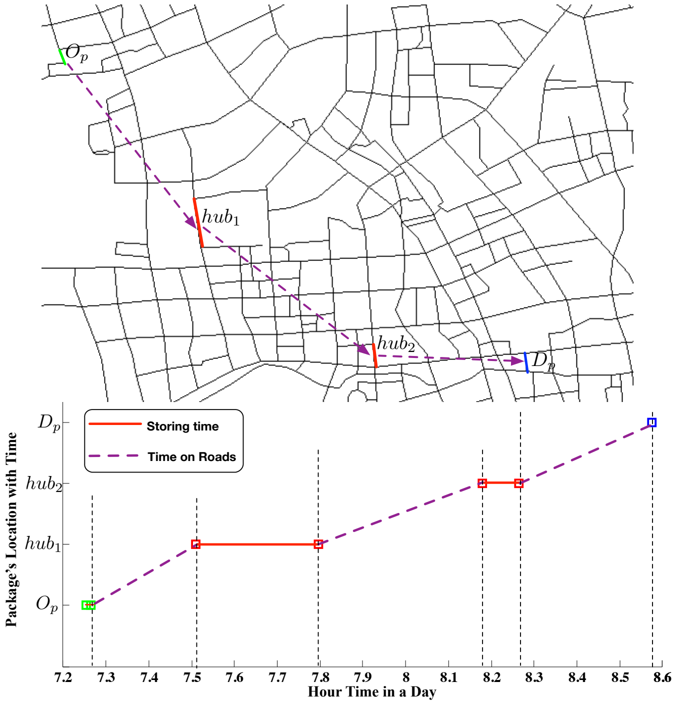

Fig. 8. Life cycle of a package from O_p to D_p with the *DesCloser* algorithm in space (top) and time (bottom). (Best viewed in the enlarged digital version)

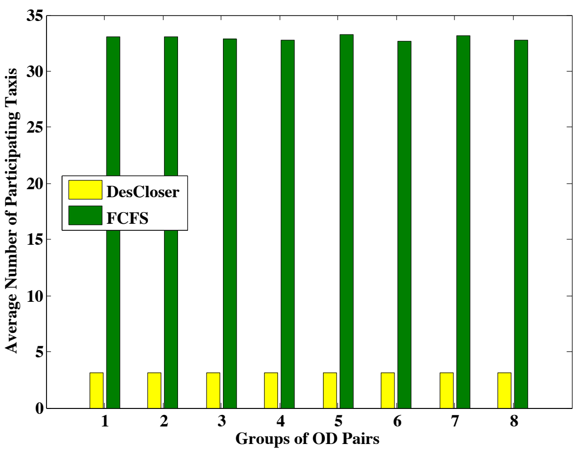

Fig. 9. Comparison result of the average number of participating taxis for two different inter-routing strategies.

We are also interested in the life cycle of the packages in space and time. We choose a package to be delivered from O_p to D_p, and plot its life cycle since its birth time in Fig. 8. To save the space, we only show the life cycle for the *DesCloser* algorithm. In this particular case, with two interchanges (i.e. hub_1 and hub_2), the package gets to its destination. As shown in the bottom figure in Fig. 8, the total package delivery time is short, taking only 1.32 hours. A taxi was recruited to collect the package and help send it to the first hub (i.e. hub_1), just 37 seconds after receiving the delivery query. Relatively more time was spent on travelling than on waiting in this case.

Fig. 9 shows the comparison of the average number of participating taxis between the two inter-routing algorithms. As expected, for the same package delivery query, the average number of participating taxis using *FCFS* (33.0) is about 10.6x greater than that with *DesCloser* (3.10).

V. Related Work

A. Crowdsourced Package Delivery

There have been a plenty of crowdsourced applications for different purposes. Among them, two relevant papers on physical package delivery relied on the spatial and time overlaps between crowdsourcing workers. Specifically, [18] involved a group of twitter users. One user passed the assigned package to another twitter user that happened to be nearby (within a certain distance). This work suffers from two main limitations: 1) it is hard to trace and coordinate the users since people rarely share their location information continuously via geo-twitter [17]. It limits the choice of suitable deliverers,probably resulting in longer package delivery delay. 2) It may be not practical to ask a participant to make a dedicated trip to pass the package to another suitable user, which may interrupt his/her on-going activities (e.g. having conversation/dinner with friends) that are hard to infer from the user's geo-twitters. The work in [13] which passed packages among mobile users with space and time overlaps based on information from cell towers had similar limitations. A spatial overlap is identified as two mobile phone users who make phones at a same cell tower which can cover a wide spatial range especially in rural areas. A time overlap is identified as the phone making time of two mobile phone users are close (within a certain time). In this work, we leverage taxis on the street which can travel longer distance. Further more, the packages are stored at the safe-boxes installed at the roads temporally between rides, and thus no time overlap is needed, requiring less participants' efforts, compared to the solutions proposed in [13], [18].

B. Taxi GPS Data Mining

Taxi GPS data has been mined for the interests of three main parties, i.e. taxi drivers, taxi passengers and city planners. Many papers recommended areas with more potential passengers to taxi drivers, e.g. [8], [23]. Zhang et al. [24] investigated the differences in three strategies between *efficient and inefficient* taxi service: passenger-searching, passenger-delivery and service-region preference. Taking a step further, based on the strategies taken by the taxi drivers, Zhang et al. [24] predicted their incomes accurately. Work on recommending the best corner to catch taxis, real-time ordering free taxis, and the taxi fee estimation targeted taxi passengers, e.g. [2], [23]. An interesting work detected anomalous taxi rides and warned the passengers "on-the-fly" when passengers [4], [5]. For city planners, taxi GPS data provides a rich source to identify flaws in city planning [25], such as determining night bus routes [6], inferring the land-use [15], etc. Castro et al. provided a good survey on understanding city dynamics from taxi GPS data [3]. In our study, taxi GPS data have been mined to identify hubs, to real-time schedule taxis to participate the package delivery tasks. Any user (taxi passengers, drivers, or ordinary residents) or company who needs package express shipping service can benefit.

C. Opportunistic Routing

Opportunistic routing is an important research topic in the delay tolerant networks (DTN), including mobile social networks, vehicular networks. In order to forward data packets from origin(s) to destination(s) under various constraints, many routing algorithms are proposed. For more details about the state-of-art routing algorithms, readers can refer to [10]. One key issue in the design of routing algorithms is to select mobile nodes which can help forward data to the destination(s). The basic idea is to select nodes which will visit the destination with high probability in the future. Many factors such as the social structure and interaction of users have been studeied comprehensively [9]. However, Chen et al. [7] argued that, in some cases, the number of nodes which can help to forward the packets is few and limited, leading to poor performance. They thus proposed an inter-landmark routing algorithm, which selects popular places (rather than nodes), and from an origin to a destination, the data packet is transited among those popular places. Our proposed framework is inspired by their idea. Unlike the data packet forwarding case in which the data can be easily copied/sent via wireless within the communication range even when the user is moving, our packages are physical and can be handed-off only at stationary safe-boxes.

VI. Conclusion and Future Work

In this paper, we have exploited the existing taxis on the street to help deliver the package while passengers are on board in a crowd-sourced manner, which is almost cost-free and environmental friendly. The work is motivated by the needs of applying pervasive sensing, communication and computing technology for *sustainable city development*. To enable the package express shipping service, we have proposed a two-phase approach. In the first phase, we first adopt a similar idea of identifying key people in the social networks to rank the road segments in the road network, according to the *influential factor*. We then locate hubs based on the ranking and their geographical locations. In the second phase, we develop two inter-hub routing algorithms to deliver the package to its destination. Finally, we evaluate the two-phase framework on a large-scale real-world taxi data set, generated by 7,600 taxis in a month. Results demonstrate that the *DesCloser* performs better than the *FCFS* algorithm in terms of the average package delivery time and the average number of participating taxis, while *FCFS* algorithm achieves higher success rate.

In the future, we plan to broaden and deepen this work in several directions. First, we attempt to develop more advanced routing algorithms to minimize the package delivery time. Second, we plan to investigate the problem with more real-life assumptions, such as the bi-directional characteristic of road network and the incentive mechanisms. Third, we would like to develop practical systems leveraging taxi GPS traces, enabling a series of pervasive smart transportation services.

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful comments and suggestions, also want to thank Nan Li for the insightful discussions. This research was partly supported by the National High Technology Research and Development Program of China (863) under Grant No. 2013AA013202, the Chongqing High-Tech Research Program under Grant No. csct2012ggC40005, and the National Natural Science Foundation of China under Grant No. 61472052, 61173014.

References

[1] Government of Hong Kong. Transport - Hong Kong: The facts, 2013.

[2] R. K. Balan, K. X. Nguyen, and L. Jiang. Real-time trip information service for a large taxi fleet. In *Proc. of MobiSys*, pages 99–112, 2011.

[3] P. S. Castro, D. Zhang, C. Chen, S. Li, and G. Pan. From taxi GPS traces to social and community dynamics: A survey. *ACM Computing Surveys*, pages 17:1–17:34, 2013.

[4] C. Chen, D. Zhang, P. Castro, N. Li, L. Sun, and S. Li. Real-time detection of anomalous taxi trajectories from GPS traces. In *Proc. of MobiQuitous*, pages 63–74, 2011.

[5] C. Chen, D. Zhang, P. Castro, N. Li, L. Sun, S. Li, and Z. Wang. iBOAT: Isolation-based online anomalous trajectory detection. *IEEE Trans. on Intelligent Transportation Systems*, 14(2):806–818, 2013.

[6] C. Chen, D. Zhang, Z.-H. Zhou, N. Li, T. Atmaca, and S. Li. B-Planner: Night bus route planning using large-scale taxi GPS traces. In *Proc. of PerCom*, pages 225–233, 2013.

[7] K. Chen and H. Shen. DTN-FLOW: Inter-landmark data flow for high-throughput routing in DTNs. *IEEE/ACM Trans. on Networking*, in press, 2014.

[8] Y. Ge, H. Xiong, A. Tuzhilin, K. Xiao, M. Gruteser, and M. Pazzani. An energy-efficient mobile recommender system. In *Proc. of ACM KDD*, pages 899–908, 2010.

[9] P. Hui, J. Crowcroft, and E. Yoneki. BUBBLE rap: Social-based forwarding in delay-tolerant networks. *IEEE Trans. on Mobile Computing*, 10(11):1576–1589, Nov 2011.

[10] B. Jedari and F. Xia. A survey on routing and data dissemination in opportunistic mobile social networks. *IEEE Communications Surveys and Tutorials*, page in press, 2014.

[11] B. Li, D. Zhang, L. Sun, C. Chen, S. Li, G. Qi, and Q. Yang. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In *Proc. of PerCom Workshop*, pages 63 –68, 2011.

[12] L. Liu, C. Andris, and C. Ratti. Uncovering cabdrivers' behavior patterns from their digital traces. *Computers, Environment and Urban Systems*, 34(6):541 – 548, 2010.

[13] J. McInerney, A. Rogers, and N. R. Jennings. Crowdsourcing physical package delivery using the existing routine mobility of a local population. In *the Orange D4D Challenge*, 2014.

[14] MIT Technology Review. Can same-day delivery succeed this time? http://www.technologyreview.com/news/520806/can-same-day-delivery-succeed-this-time/, 2014.

[15] G. Pan, G. Qi, Z. Wu, D. Zhang, and S. Li. Land-use classification using taxi GPS traces. *IEEE Trans. on Intelligent Transportation Systems*, 14(1):113–123, 2013.

[16] F. Research. E-retail spending increase 45% by 2016. http://www.internetretailer.com/, 2012.

[17] A. Sadilek, H. Kautz, and J. P. Bigham. Finding your friends and following them to where you are. In *Proc. ICWSM*, 2012.

[18] A. Sadilek, J. Krumm, and E. Horvitz. Crowdphysics: Planned and opportunistic crowdsourcing for physical tasks. In *Proc. ICWSM*, 2013.

[19] J. Shetty and J. Adibi. Discovering important nodes through graph entropy the case of enron email database. In *Proc. of LinkKDD*, pages 74–81, 2005.

[20] The Economist. Same-day dreamer. http://www.economist.com/news/business/21601556, 2014.

[21] M. Veloso, S. Phithakkitnukoon, and C. Bento. Urban mobility study using taxi traces. In *Proc. of TDMA*, pages 23–30, 2011.

[22] J. Yuan, Y. Zheng, X. Xie, and G. Sun. T-drive: Enhancing driving directions with taxi drivers' intelligence. *IEEE Trans. on Knowledge and Data Engineering*, 25(1):220–232, 2013.

[23] N. J. Yuan, Y. Zheng, L. Zhang, and X. Xie. T-finder: A recommender system for finding passengers and vacant taxis. *IEEE Trans. on Knowledge and Data Engineering*, 25(10):2390–2403, 2013.

[24] D. Zhang, L. Sun, B. Li, C. Chen, G. Pan, S. Li, and Z. Wu. Understanding taxi service strategies from taxi GPS traces. *IEEE Trans. on Intelligent Transportation Systems*, in press, 2014.

[25] Y. Zheng, Y. Liu, J. Yuan, and X. Xie. Urban computing with taxicabs. In *Proc. of UbiComp*, pages 89–98, 2011.