A crowdsourcing solution to collect e-commerce reverse flows in metropolitan areas

Shenle Pan, Chao Chen, Ray Y. Zhong

A crowdsourcing solution to collect e-commerce reverse flows in metropolitan areas

Shenle Pan*, Chao Chen, Ray Y. Zhong*****

** MINES ParisTech, PSL Research University, CGS - Centre de gestion scientifique, 60 Bd St Michel 75006 Paris, France (Tel: +33 1 40 51 93 12; e-mail shenle.pan@mines-paristech.fr).*
***School of Computer Science, Chongqing University, (e-mail:cschaochen@cqu.edu.cn)*
**** Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, (e-mail: zhongzry@gmail.com)*

Abstract: On the forward side, the growth of E-commerce in recent years substantially generates additional packets and parcels for distribution; meanwhile, on the reverse side, collecting returned goods is also becoming a preoccupation of sustainability, especially in metropolitan areas. Inspired by the concepts of *crowdsourcing* and the *Physical Internet*, in this paper, we propose an innovative solution that seeks to exploit the extra loading capacity and constant mobility from taxis in metropolitan areas to collect and delivery the e-commerce returns from final consumption points back to retailers. We assume that, on one hand, e-retailers will have incentive to outsource this task; on the other hand, taxi drivers will also be motivated because they can earn a little extra money from the shipments that they have fulfilled. As an alternative to the traditional ways, the solution proposed is more sustainable because it could simultaneously reduce the economical (pick-up and transportation costs), environmental (CO_2 emissions, energy consumption, traffic congestion in city), and social (the wastes of the impulse buying, reduced incitation of online shopping) impacts resulted from reverse flows management in metropolitan areas. As the first qualitative and quantitative study of the concept, this paper uses open databases of taxi GPS traces and locations of shops in a large city in China for investigating the feasibility and viability of the solution proposed. Two collection strategies are proposed and evaluated by an optimization-based simulation model. The results generate several useful insights to the implementability and managerial issues of the concept.

Keywords: E-commerce reverse logistics, Crowdsourcing, Physical Internet, Collection problem.

1. INTRODUCTION

Managing reverse flows is becoming an important issue in e-commerce logistics (Fleischmann et al., 1997; Rogers et al., 1999). For example, on the forward side, the largest e-commerce company *Alibaba Group* just announced that, on the day of 11th November 2014, the transaction in 24 hours via their platform was 57 billion RMB (which is 9,3 billion USD) resulted from 278 million on-line purchasing orders, which is a new record against 35 billion RMB of 180 million purchasing orders in 2013 (www.alibabagroup.com). However, little attention has been paid to the reverse flows: the percentage of returned goods is in average 25%, even 40% for some products like apparel, out of the on-line purchased goods on the same day in 2013 (source: www.sina.com). Considering the importance of e-commerce reverse flows, we can infer that the returns collection problem is becoming a more and more notable issue of sustainable development. This problem is particularly observable in metropolitan areas, not only due to the economical preoccupations related to the pick-up costs, but also because of the environmental footprints (CO_2 emissions, energy consumption, traffic congestion) and the social impacts (the wastes of the impulse buying, reducing the incitation of online shopping). And comparing to the distribution of forward flows, the collection problem of reverse flows has its own characteristics: low added value, the same destination (for items from the same retailer), flexible delivery time, less important protection, etc. In this context, this paper seeks to propose a sustainable solution to the e-commerce returns collection problem.

In the literature, the retail (or e-retail) returns collection problem is generally defined as to collect and transport return goods from consumption points to distribution centers of retailers (Wojanowski et al., 2007). It has been mainly studied in two streams: the collection-network design (Bostel et al., 2005; Wojanowski et al., 2007); and the vehicle routing optimization in city (Beullens et al., 2004). For the sake of sustainable development, this paper proposes an alternative solution conjointly inspired by the concepts of crowdsourcing transport (Estellés-Arolas & González-Ladrón-de-Guevara, 2012; McInerney et al., 2014; Sadilek et al., 2013), and the Physical Internet (Ballot et al., 2014; Mervis, 2014). The idea is based on a principle that is to crowdsource the task of collecting return goods in metropolitan areas to people who are willing to earn a little extra money. As an illustration, this paper investigates the idea with application to taxi drivers for two major reasons: their constant mobility that mostly covers the shipments paths; and their movement traceability via GPS that ensures the monitoring and the traceability of goods. One the other hand, using extra loading/mobility capacity to delivery goods for earning little money could also be attractive to taxi drivers, while certainly without compromising the service to their clients. The paper contributes to the quantitative researches investigating the feasibility of crowdsourcing solution to the retail returns collection problem in metropolitan areas. Moreover, it also contributes to the undergoing practical projects of crowdsourcing last-mile delivery, e.g., projects in Amazon[1].

This paper focuses on addressing two questions. The first is how to design a sustainable collection network that is flexible and efficient to taxi drivers and the second is how to propose efficient and reliable collection strategies using the network. To answer the first question, we propose to use shops as collection infrastructures as discussed in (Beullens et al., 2004). Because they can provide advantages like flexible drop-off and pick-up time, high accessibility to consumers and drivers, as well as some necessary controls on returned goods. It means that the generators may drop their packages at the nearest shops without planning and waiting for an on site pickup service. Additionally, shops have also high volume of returned goods for example the surplus stocks, outmoded clothes etc. Hence, we need to define a collection network that consists of not all but the most reliable shops to taxi trajectories. Then, appropriate collection strategies are necessary to allow taxis to efficiently transport returned goods from shops to distribution centers. Two strategies are studied in this paper. They are DIRECT and ROUTING. The DIRECT means goods should be transported by only one taxi from source to destination, while ROUTING allows transshipment between taxis.

This paper is organized as follows. The next section will present the main assumptions and definitions of the problem. In Section 3, two collection strategies are defined and investigated. Section 4 is concerning with an experimental study based on an application to a large city in China with real data of shops location and taxi GPS trace. The simulation results will also be discussed. Finally, in Section 5 we will conclude our works and give some perspectives of future work.

2. PROBLEM DEFINITION AND ASSUMPTIONS

2.1 Assumptions

Some assumptions are proposed in this study so as to respect some practical constraints and to reduce the complexity of the problem.

Assumption 1. Return goods, hereafter packages, can only be assigned to taxis with passenger. At this first stage we investigate the possibility of using extra loading capacity of taxi without challenging their current clients hunting strategy.

Assumption 2. Delivering package should not have impact on the service to passenger, e.g., Passenger service first-Package service second. In other words, drivers may not change the route to pick-up or drop-off a package when passengers are on taxis. The Assumptions 1 and 2 aim also to limit the environmental footprint of the proposed solution: fulfilling package delivery services will barely degenerate extra energy consumption thus CO_2 emission since it does not change taxis' route.

Assumption 3. All operational constraints are not considered at this stage, i.e., taxis' loading capacity, shops' storage capacity and working time, road and traffic condition.

Assumption 4. All taxi drivers and shops in the city will accept the proposed solution. This paper focuses on feasibility study of the solution. Business models or plans to drivers or shops are not considered here.

2.2 Definitions

Road Network. A road network (connection-network) is a directed graph that is the road transportation network for physical distribution in a city. As illustrated in Fig. **1**, each edge connects two nodes (circles), and is a bi-directional edge since it usually has two driving directions. We particularly denote e_{i+} as the right-hand driving direction, while e_{i-} as the other driving direction for edge e_i.

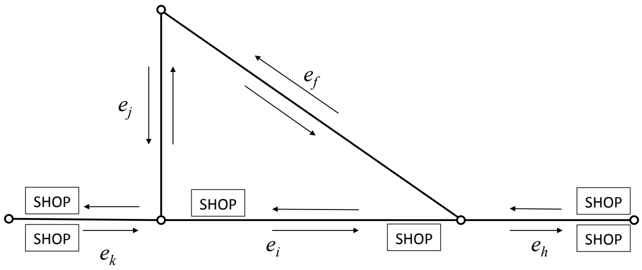

Fig. 1. Illustration of a collection network in city level.

Link Graph. A link graph is a directed graph defined as $G=(V,E)$, where V is the set of road sections with respect to direction (lines with arrow), i.e., edges of the road network noted as $\{e_{f\pm}, e_{i\pm}, e_{j\pm}, e_{k\pm}, e_{h\pm}\}$ in Fig. **1**; and E is the set of directed edges that are taxi passenger flows from source node to destination node, for example $e_{h+} \rightarrow e_{k+}$, and noted as E_{h+k+}.

Collection facilities. Shops are the collection facilities considered in this paper. We assume that each road section has at most two shops (for different driving directions), as shown in Fig. **1**, to reduce the complexity of the network. A shop is also a POI (Point of Interest) in the network.

Hotlines. We define here the hotlines as the top-k frequent edges in link graph (G) computed by the accumulation of taxi passenger flows. In other words, only the top-ranked edges in a link graph will be selected and maintained in the network (the top 100 for example). It is denoted as h_{i+j+} if E_{i+j+} is a hotline.

Package service and passenger service. This paper considers package delivery and passenger delivery. Their service request is defined by a triplet <origin, destination, birth time>. We assume that the delivery time windows for

[1] http://www.engadget.com/2014/11/05/amazon-is-exploring-taxi-deliveries-in-san-francisco-and-los-ang/

reverse flows are flexible so that the arrival time is not considered as a constraint. Accordingly, a package service and passenger service placed are denoted as <o_p, d_p, t_p> and <o_t, d_t, t_t>.

3. COLLECTION STRATEGY

As pre-defined, road network and link graph are respectively representing the physical and flows network. The first one is for computing the location of package and distance of delivery, while the second one is for matching the package flows and passenger flows for simulation. The next step is to define appropriate collection strategies for simulation. Two strategies are proposed and detailed presented as follows.

3.1 Strategy DIRECT

Intuitively, the simplest way is that a package waits in a shop for a passing taxi having exactly the same destination (i.e., the edge of consolidation center); and no transshipment between taxis is allowed. It is called DIRECT strategy. The algorithm for this strategy is as follow: for a given package request <o_p, d_p, t_p>, after occurrence it will be picked-up by the first taxi <o_t, d_t, t_t> which has the same destination, i.e., $t_p \leq t_t$ and $d_p = d_t$. There are some limitations in this strategy. First, packages may not be collected if there is no direct taxi trajectory for the occurring point to destination. Second, the efficiency of time or transport may be low. It is hence a baseline scenario to the experimental study.

3.2 Strategy ROUTING

If we assume that transshipment between taxis are allowed, which means that a package can be delivered by one or several taxis in sequence to reach its destination, the waiting time could be improved as well as total delivery distance. Then the problem is similar to the container routing problem in Physical Internet studied in (Sarraj et al., 2013). So it is called ROUTING strategy. However, the transfer times should be limited for the reason of handling costs and the risk of damage to a package. For a given package request <o_p, d_p, t_p>, its routing path is determined by 3 steps:

Step 1. Identify all hotlines between o_p and d_p in the link graph.

Step 2. Select the hotlines having *compatible* direction with the o_p-d_p direction. We first keep hotlines lying between the range of [$o_{p \cdot x\text{-}thx}$, $d_{p \cdot x+thx}$], and [$o_{p \cdot y\text{-}thy}$, $d_{p \cdot y+thy}$]; then select hotlines in the range whose direction is compatible with that of op-dp, that is, the direction of selected hotline should have an angle less than 90° with the direction of o_p-d_p: Angle (o_p-d_p, h_{ij})<90°. For example, in Fig. **2** the hotline h_{jk} will not be selected since Angle (o_p-d_p, h_{jk})>90°.

Step 3. From the set of hotlines selected, to limit the number of transshipments, we simply identify the most frequent one as the relay path, thus for a package from o_p to d_p, it will be first shipped to origin of the selected hotline, and then shipped to the destination of the selected hotline, finally shipped to the destination of the package from the destination of hotline, that it, (o_p→*hotline.o*→*hotline.d*→d_p). Note that if, occasionally, the h_{od} itself is the most frequent hotline, then it will be a direct route. Otherwise, if the route consists of several intermediate hotlines, they can be seen as hubs of Physical Internet. For example, in Fig. **2** the *Route*(h_{oi}, h_{ik}, h_{kd}) is the optimal from *o* to *d*; and h_{ik} is actually an hub for transshipment.

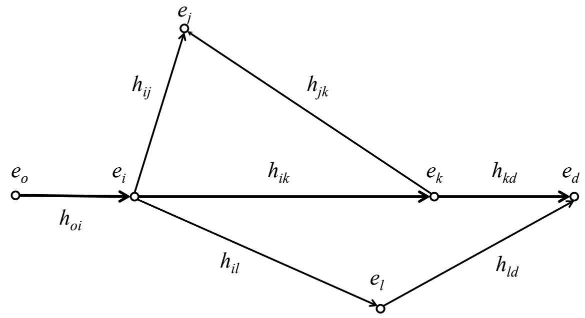

Fig. 2. Example of hotlines for package from e_o to e_d in a link graph

Once the optimal route is given to a package, it should wait for the first passing taxi driving it to the origin of the identified hotline or directly to its final destination if the later is optimal, equally to the next section if there has. Comparing to the DIRECT strategy, this strategy is more flexible and that could improve the efficiency of time and transport.

4. EXPERIMENTAL STUDY

4.1 Methodology and Input Data

This part descripts an experimental study conducted for evaluating the proposed solution. We use an open database of taxi's GPS trace for a simulation study. The data source is from Hangzhou, which is a large city in China for carrying out the experiments.

Firstly, to design a collection network, we have located from *Baidu Map* more than 3000 "shop type" POIs distributing the whole Hangzhou city (area of 15km*30km). And only around 852 shops are kept as we maximally selected 2 shops in each road section (one for each direction). Arrows in Fig. **3** indicate the driving directions of the corresponding shops. As an assumption, four POIs at the four corners of the map are selected as distribution centers, one of whom all packages should go to.

Fig. 3. 500 shops with road direction located in Hangzhou city (x=latitude and y=longitude) (Best viewed in the enlarged version)

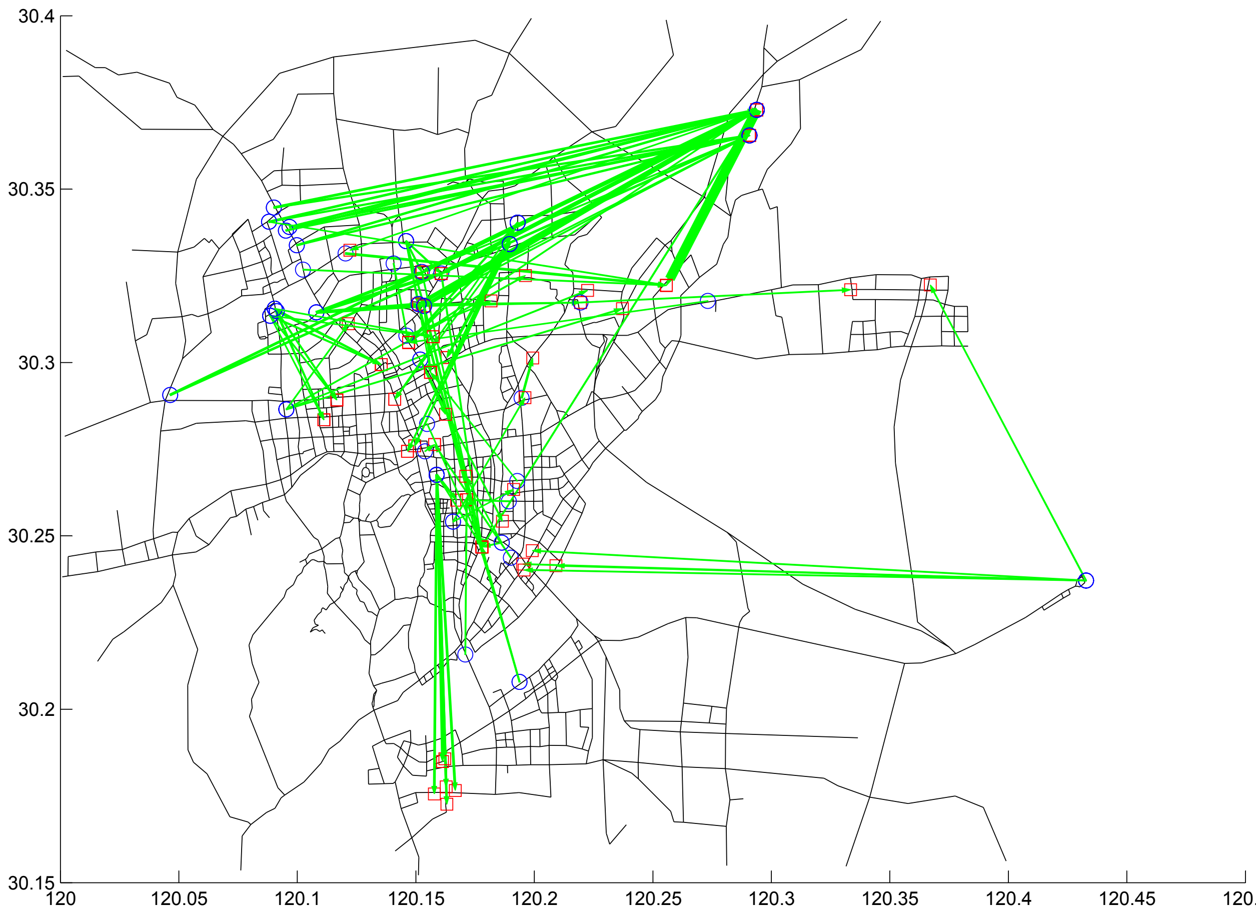

Fig. 4. Hotlines in the link graph (top-100 ranked edges with x=latitude and y=longitude, circles=origins and squares=destinations) (Best viewed in the enlarged version)

Second, based on historical taxi GPS data (e.g., one-month), each element in the link graph (G) can be derived, as taxi GPS data can tell us the passenger flow information from which road section to which road section at what time. The detailed taxi GPS data description can be found in our previous work (Castro et al., 2013; Chao et al., 2014). Afterwards, we can identify the hotlines in the link graph by simply ranking. Fig. **4** shows the result of top-100 hotlines in the link graph based on one-month (January in 2010) taxi GPS data. Due to the complexity of computation, only the top 100 hotlines are selected in this paper. But we believe that these hotlines already cover the active districts of Hangzhou city.

Third, as the time horizon of simulation is fixed for a month, 2 000 packages are randomly generated (within the month) and each package has its proper value to the triplet $<o_p, d_p, t_p>$. With regard to the complexity, only few packages are generated and simulated in the model comparing to the reality. However, we believe that the quantity can still yield reliable simulation results.

Table 1 summarizes input data to the study.

Table 1. Input data (#: number)

Time (H)	Area (km*km)	# Shops	# DC	# Packages	# Taxis
720	15*30	852	4	2 000	>7 000

4.2 Results

The model is coded in Matlab and run on an Intel Quad CPU PC with 12G RAM and Windows 7 operation system. Two scenarios are simulated and for each package two KPI (key performance indicators) are computed: the delivery time and the total transportation distance in km. Without giving all details of the result, Fig. **5** and Fig. **6** illustrate the CDF (cumulative distribution function) of the two KPIs.

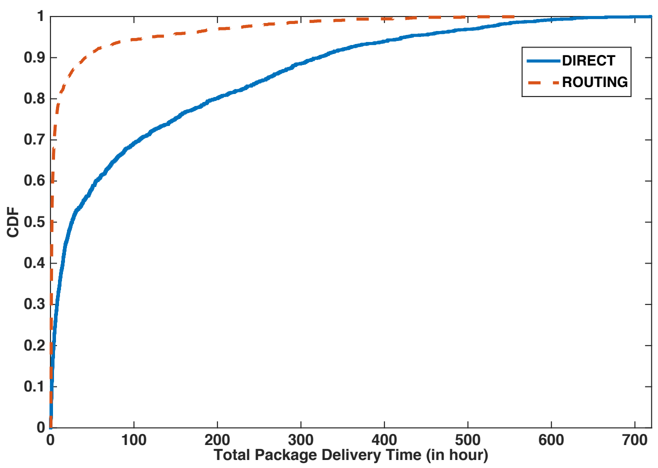

Fig. 5. The cumulative distribution function of the delivery time (time horizon=720 hours)

As seen in Fig. **5**, the ROUTING strategy obviously performs better in delivery time: almost 90% packages are fulfilled within 24h with ROUTING strategy, while only around 55% with DIRECT strategy. In addition, only few packages, around 5%, in ROUTING strategy have delivery time longer than 100h, that is around 4 days. We believe that this is because the packages are randomly generated so that some of them may not be well covered by the hotlines and the schedule of the taxi fleet.

Fig. 6. The cumulative distribution function of the transportation distance (area of 15km*30 km)

According to Fig. **6**, the ROUITNG strategy may generate roughly 10 km more per package than the DIRECT strategy, simply because the routing path may be longer than the direct path. However, we recall that both of the two strategies have the same assumptions that will not change the service to passengers and the schedule of the taxi fleet. That means the extra distance will barely generate any extra environmental footprint.

In the next step a scenario of collaborative urban consolidation center (Faure et al., 2014) to robust the solution proposed will be tested. It means a single center will be defined as the final destination for all packages. There are several reasons to study this scenario. First, a collaborative center will probably mitigate the managerial issues of the solution proposed (reception at distribution center, handling etc.); second, using the center as a unique final destination can better consolidate the reverse flows as well as the forward flows in the city, then further robust the solution proposed; third, if the consolidation center is well located, some of the KPIs could be improved (i.e. delivery time, successful delivery rate, distance). The expected benefits will be quantitatively studied in the next steps.

5. CONCLUSION

It has been years since the word "crowdsourcing" was firstly mentioned in the industry (Howe, 2006). However, only few studies can be found from the literature, even less on freight transportation of reverse logistics in city[2]. To fill the gap, this paper studied an innovative alternative solution to e-commerce returns collection problem in metropolitan areas, which uses shops and taxi drivers in the city for collecting and transporting returns to retailers. An experimental study, with real data of shops' location and taxi traces from a large city in China, has been conducted to analyze the feasibility of

[2]http://en.wikipedia.org/wiki/List_of_crowdsourcing_projects

the solution. Although it is difficult to directly compare the solution proposed with the real-word cases that have very different collection strategies, we can still claim that the solution proposed is more sustainable. First, since we aim to exploit extra capacity of taxis in city without changing their current strategies, the solution will barely generate extra environmental footprint when considerably eliminating environmental problems from the trucks for the same purpose. Second, it is believed that the crowdsourcing solution proposed will be more economic than the current practices, as proved in some other cases (Brabham, 2008) and (Singer & Mittal, 2013). This question will be further studied in the next steps. Third, the solution is also social-friendly because it gives chances to taxi drivers to earn a little extra money and it offers the final consumers an alternative and easier way to return the unwanted products. Moreover, since the solution proposed has been investigated with real data of shops' location and taxi traces, and with delivery demands randomly generated in the city, the experimental study is very close to a direct application and the results obtained are significant to the industry.

This paper only initializes the study with limited data samples. The future works could be carried out from several aspects. For example, we can extend the network from shops to other point of interest like post office, automated lockers, even taxi stations etc. Additionally, the solution can also be generated to taxi flows without passenger, organizations like Uber, even private cars. Other then the feasibility question addressed in this paper, the motivation of the participants (drivers or shops) and their decision should also be studied. To this end, some researchers have paid attention to the pricing mechanism problem in crowdsourcing market (Singer & Mittal, 2013), while others focus on the business models (Brabham, 2008). At last, the traceability, standard size and protection of packages are also related to crowdsourcing solution. One of the solutions is to use intelligent modular containers instead of today's cardboard boxes, as proposed in (Ballot et al., 2014) and (Sallez et al., 2014). The container intelligence can not only help trace and optimize the packages' route, but also collect real time transportation information to be used to further analyses.

REFERENCES

Ballot, E., Montreuil, B. & Meller, R. (2014). The Physical Internet: The Network of Logistics Networks: La documentation Française.

Beullens, P., Van Oudheusden, D. & Van Wassenhove, L. (2004). Collection and Vehicle Routing Issues in Reverse Logistics. In: R. Dekker, M. Fleischmann, K. Inderfurth & L. Van Wassenhove, Reverse Logistics (pp. 95-134): Springer Berlin Heidelberg.

Bostel, N., Dejax, P. & Lu, Z. (2005). The Design, Planning, and Optimization of Reverse Logistics Networks. In: A. Langevin & D. Riopel, Logistics Systems: Design and Optimization (pp. 171-212): Springer US.

Brabham, D.C. (2008). Crowdsourcing as a model for problem solving an introduction and cases. Convergence: the international journal of research into new media technologies, 14(1), 75-90.

Castro, P.S., Zhang, D., Chen, C., Li, S. & Pan, G. (2013). From taxi GPS traces to social and community dynamics: A survey. ACM Computing Surveys (CSUR), 46(2), 17.

Chao, C., Daqing, Z., Nan, L. & Zhi-Hua, Z. (2014). B-Planner: Planning Bidirectional Night Bus Routes Using Large-Scale Taxi GPS Traces. Intelligent Transportation Systems, IEEE Transactions on, 15(4), 1451-1465.

Estellés-Arolas, E. & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information science, 38(2), 189-200.

Faure, L., Montreuil, B., Marquès, G. & Burlat, P. (2014). A basic collaborative city logistics' solution: the Urban Consolidation Center. In: 7th International Conference on Interoperability for Entreprises Systems and Applications (I-ESA 2014). Albi, France.

Fleischmann, M., Bloemhof-Ruwaard, J.M., Dekker, R., van der Laan, E., van Nunen, J.A.E.E. & Van Wassenhove, L.N. (1997). Quantitative models for reverse logistics: A review. European Journal of Operational Research, 103(1), 1-17.

Howe, J. (2006). The rise of crowdsourcing. Wired magazine, 14(6), 1-4.

McInerney, J., Rogers, A. & Jennings, N.R. (2014). Crowdsourcing Physical Package Delivery Using the Existing Routine Mobility of a Local Population.

Mervis, J. (2014). The information highway gets physical: The Physical Internet would move goods the way its namesake moves data. In: Science (Vol. 344, pp. 1057-1196).

Rogers, D.S., Tibben-Lembke, R.S. & Council, R.L.E. (1999). Going backwards: reverse logistics trends and practices (Vol. 2): Reverse Logistics Executive Council Pittsburgh, PA.

Sadilek, A., Krumm, J. & Horvitz, E. (2013). Crowdphysics: Planned and Opportunistic Crowdsourcing for Physical Tasks. SEA, 21(10,424), 125,620.

Sallez, Y., Montreuil, B. & Ballot, E. (2014). On the Activeness of Physical Internet containers. In: T. Borangiu, A. Thomas & D. Trentesaux, Service Orientation in Holonic and Multi-Agent Manufacturing Control (pp. 235-244): Springer Berlin Heidelberg.

Sarraj, R., Ballot, E., Pan, S., Hakimi, D. & Montreuil, B. (2013). Interconnected logistic networks and protocols: simulation-based efficiency assessment. International Journal of Production Research, (In Press).

Singer, Y. & Mittal, M. (2013). Pricing mechanisms for crowdsourcing markets. In: Proceedings of the 22nd international conference on World Wide Web (pp. 1157-1166): International World Wide Web Conferences Steering Committee.

Wojanowski, R., Verter, V. & Boyaci, T. (2007). Retail–collection network design under deposit–refund. Computers & Operations Research, 34(2), 324-345.