EPiC
Computing

A Novel Delivery Model to Expand E-commerce Customers Based on Telecom Data Mining

Giridhar Maji[1], Sharmistha Mandal[2], Narayan C. Debnath[3], and Soumya Sen[4]

[1] Department of Electrical Engineering
Asansol Polytechnic, Asansol, India
giridhar.maji@gmail.com
[2] Department of Computer Science & Engineering
University of Calcutta, Kolkata, India
sharmistha.cse@gmail.com
[3] Department of Software Engineering
Eastern International University, Vietnam
narayan.debnath@eiu.edu.vn
[4] A.K. Choudhury School of IT
University of Calcutta, Kolkata, India
iamsoumyasen@gmail.com

Abstract

A very important issue with the e-commerce delivery service in most of the emerging economies including India is the last mile connectivity. Delivering products, booked online to the remote tier-2 and tier-3 cities remained "costly". It is observed from firsthand experience with some well-known e-commerce brands in India that their delivery service partners tend to cancel orders that are far away from their tier-2 logistics hubs with the reason shown as "address out of delivery range". Due to low order density in the far flanges of tier-2 and tier-3 cities arranging vehicles and delivery personnel become costly. In this paper, we propose an innovative delivery model to serve the remote areas by opening edge-hubs at selected places and employing local daily commuters for last mile delivery. Identifying the edge-hubs for opening distribution centers is a costly business if done using traditional field surveys. Here we propose the use of telecom call detail record (CDR) location data as an alternate way of identifying the hubs in real time with much less cost and time.

1 Introduction

The mobile phone has saturated the global population. Nowadays almost every person has got a mobile phone. For example, India is the 3rd largest market of the smartphone with around 125 Million users. Again smartphone use is also increasing rapidly with the reduction of price and introduction of 3G and 4G in developing countries [13]. As an example, India ranked 2nd depending on the total number of Internet users from any device. Let us focus specifically on India as a representative developing country. With the increasing mobile network penetration

to rural areas and state initiatives like 'Digital India'[2] has fueled the Internet use by mobile net or broadband in greater suburban areas including tier-2 & tier-3 cities. In India, 1.5 GB/day of prepaid mobile data is available for a validity period of 4 weeks with a mere 2 USD bill. Hence, the young generation of India (and most other developing countries) using their smartphone with cheap data plans for everything digitally possible. Popularity, as well as market share of any e-commerce business, mainly depends on the infrastructures such as internet availability, smartphone use, low-cost mobile data plans, and broadband services in suburban areas. As per the current statistics of the Indian retail e-commerce business turnover [15] shown in figure 1, it is apparent that with the growth of smartphones and cheap data plans online retail business is flourishing coherently. Smartphone and Internet users in suburban and rural areas are using the internet more and more for everything. Initially, they start with social networking sites like facebook, twitter etc. and finally start venturing into the world of e-commerce websites or e-retailer apps. A few popular such e-commerce companies operating in India include Amazon, Flipkart, Snapdeal, Alibaba etc. With the advent of "Cash on demand (COD)" feature by the e-commerce businesses even requirement of having an online banking account or debit or credit card to make a purchase is removed. Also due to many government initiatives (Pradhan Mantri Jan Dhan Yojana [7]), almost every Indian family now has a bank account with a Rupay card [5], which can be used for such payments. As a result, e-commerce businesses are continuously increasing their market share against their retail store counterpart.
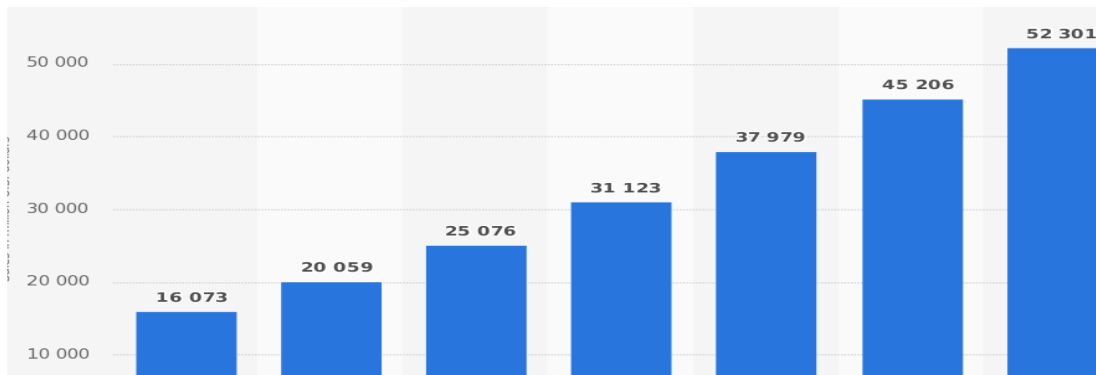

Figure 1: e-commerce sales in India in million USD projected up to the year 2022

Even then yet market share of e-commerce companies is much less than that of their counterpart due to the reason that e-commerce delivery services has not yet matched the pace of the Internet penetration and is limited to only tier 1 and in some cases tier 2 cities within a very small periphery. E-commerce delivery partners generally do not take delivery orders for the suburbs because of 'not being profitable'. This is mainly due to the reasons that order density i.e. number of orders to an area is less and cost to deliver is more per delivered item. We will try to solve the problem of this last mile connectivity- from tier 2 city boundary to the nearby suburban areas with some innovative and novel proposal. First with the help of data mining on telecom call detail record (CDR) data we shall identify important tier-2 and tier-3 cities which are well connected by many daily commuters to the nearby suburbs and rural areas. It is those areas from where almost every day a lot of people travel to the tier-2 or tier-3 cities. Now if e-commerce delivery has services up to the tier-2 or tier-3 city then they will be able to cater to the nearby areas with our unique business proposal.

Rest of the paper is organized as follows. Problem statement and motivation to consider

this problem has been presented in section 2. Background study on telecom CDR along with discussion on related e-commerce delivery models are done in section 3 followed by the proposed delivery model in section 4 and methodology followed in section 5 that details the CDR data preprocessing and proposed method to identify delivery hubs. Finally section 6 concludes the paper.

2   Problem Statement and Motivation

Author has once ordered an item from one of the online electronic stores using COD feature. Author lives in an area which is around 8 km from nearby tier-2 city. The e-commerce delivery service provider has their delivery hub at that city. The ordered item reached to that city from central hub in 2/3 days and then the order gets cancelled as "the address is out of delivery range". The author then made some enquery with people from neighboring areas and found out that a lot of them have faced such kind of problem. It is understood that even in rural areas there are lot of people with the internet access and ready to purchase online but turned down due to the delivery bottleneck. The author came to know that either those people try and arrange delivery with some friend's place in cities or goes to the retail store for shopping. This incident motivates authors to analyze the situation and try to find a win-win solution that will make both the parties happy. Kshetri [6] has done a case study to understand the barriers to e-commerce in developing countries. He has identified 3 types of barriers such as, (i) economic (ii) Socio-political and (iii) cognitive. In granular level some of the important barriers include (i)availability of websites and awareness among customers (ii) ICT skills of the customers (iii) low cost internet availability (iv) web contents in local language and trust on the e-commerce company (v) strict law to protect from online fraud (vi) delivery infrastructure and (vii) delivery service reachability etc. As discussed in the introduction, in most of the developing economies the main barrier that remains after widespread cheap Internet and smart phones is the delivery service to the remote parts of the country. The objectives of the paper are:

- to propose a delivery model that will overcome the delivery service reachability problem to the sub urban areas near to a tier-2 or tier-3 city in such a way that both the parties i.e. the customer and the online seller gain from the solution.

- to identify the geographic locations to open edge-hubs/delivery points using telecom data mining.

3   Background Study And Related Work

Telecom Call Detail Records (CDR) has recently been profoundly used for various business applications such as (i) understanding tourist movement patterns in a city [1],(ii) to understand the market share of a mobile handset maker [10], (iii) identifying important places in a city [4], (iv) predicting churn in telecom customers [11] etc. We shall be employing data mining on telecom call details record (CDR) data to identify target cities where extra delivery kiosk can be open for our proposed business solution. To understand the process of identifying the cities those are strongly connected (by a lot of daily commuters from nearby suburbs) to a large number of nearby areas. Let us first understands what CDR data is and how it is used to estimate a person's location and his daily commuting pattern in the next sub sections below.

3.1   Telecom CDR

Whenever a mobile call is placed it follows certain protocols to reach the destination user from the caller. A call can only be made if the calling party (his cellphone) is under the network coverage of some BTS (base transceiver station) connected to a cell tower antenna by the telecom network service provider. Each mobile network tower antenna creates a cell within it's radio link reachability. Each such cell has a unique number associated known as CellID. The CellID provides a low-cost user location estimation (GPS provides much finer location but that incurs extra power and not all basic mobile devices are equipped with GPS system) based on the location of cell tower to which a user is connected. When ever a user moves out of one cell to another new cellid is updated in SIM(Subscriber Identity Module) as well as in CDR. A typical entry of telecom CDR contains many technical parameters including billing details. We shall only consider only limited number of fields as outlined in table 1 from CDR database that are important for our specific objective. Here calling party number or called party numbers could be replaced by any unique codes to preserve the user privacy as we are mainly interested in the location of the users to identify the aggregated movement pattern.

Table 1: Few sample CDR fields used in the study.

Caller ID	Caller CellId	Caller LAC	Callee ID	Callee CellId	Callee LAC	call start time	duration
25358XXX	5102	41708	25358XXX	3981	56292	2018-12-06 14:24:32	00:15:34

3.2   E-commerce Delivery Models

One of the important difference between online e-commerce market place and brick-and-mortar shops down the street is the mode of product delivery. In case of brick-and-mortar shop or large retail shops customers normally chooses the products and then pay using card or cash and then go away with the brought items. A typical work-flow at an Point-of-Sale counter in retail stores are discussed in [9]. In case of e-commerce, products are displayed on web portals where customer chooses and pay either using card or by cash-on-delivery(COD [3], a feature Indian e-commerce companies are employing to gain trust of the first generation Internet users). Now the problem of delivering the products and goods to the customer premises and handing them over to receive payment for COD orders has become a larger issue as it involves delivery vehicle route planning, customer availability at home, preferred time window etc. along with profitability. Good survey of different delivery models and related logistics issues are considered in [8, 16]. Reception box and delivery box [14] are two popular e-grocery delivery models where in the first case each customer is given a secure reception box with refrigeration facility and delivery person put the ordered items inside that can only be opened by the customer upon returning home. In the latter model a delivery box is delivered with the products in the basement or garage of the customer and on the next day, empty box is taken back by the delivery vehicle. In both the above models high cost is involved. Reception box could only be viable for the regular customers with high volume of orders. In India most of the customers are just starting to use online services and home deliveries. So above models do not work. Here we propose a new low cost delivery model using local people without incurring any delay in item delivery.

4   Proposed Delivery Model

We propose to establish edge-hubs at the tier-2 and tier-3 cities that will act as delivery points for the nearby sub urban areas and also rural areas. Local daily commuters from the surrounding rural areas will register themselves as a part-time delivery agent at the kiosk/delivery point. Some amount of caution money could be taken to ensure delivery of product. Now in the afternoon when those people returns from work they will pick up the items that are to be delivered in their neighborhood places and then in the evening they will deliver the items to the customers. In present condition delivery agents are given fixed fuel cost and commission per item for delivering items. Due to low order density and longer travel between one customer location to another delivery agents are not interested in delivering to the remote areas, instead they focus on serving within city boundaries where they could delivery higher number of items with less travel and in less time to earn more. In the proposed delivery model commuters will use their own vehicle to deliver in their nearby places. Small commission per item will be a big addon to their small daily earning. Thus, this model not only serves the e-commerce logistics companies with real 'pan India' delivery with no extra cost, at the same time it enables in the social upliftment of the poor.

Identifying the edge-hubs for opening distribution centers is a costly business if done using traditional field surveys. Here in the next subsection we propose a method to identify edge-hubs with the help of CDR data mining.

5   Methodology

Any analysis on CDR data for spatial application should be preceded with an integration of CDR data from many telecom operators [12] of the concerned area to get a fairly correct estimate. Though even with data from the top telecom operator with highest market share in that area might give quite reasonable outcome.
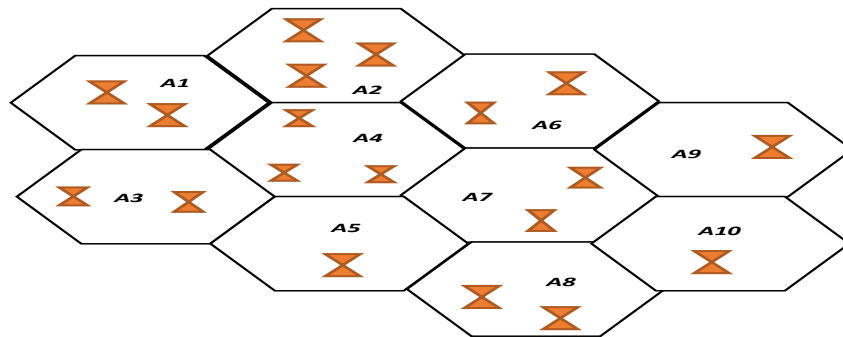

Figure 2: City surrounding areas divided into small sub-areas with cell towers.

5.1   Data Preprocessing

1. Divide the selected surrounding sub urban area of tier-2 and tier-3 cities into smaller regions that are covered under a few cell towers as shown in figure 2. Area $A1$ consists of all area under the 2 cell towers. Similarly, Area $A4$ is covered by 5 cell towers. It is quite appropriate to consult local geographic map while dividing the region into smaller areas.

Table 2: Intermediate data preprocessing

Unique UserID	Home Area	Work area	Hour_Stayed
USER001	A1	A7	3
USER002	A2	A4	4
USER003	A2	A4	5
USER004	A3	A4	7
USER005	A7	A5	1
USER006	A10	A4	9
USER007	A8	A2	4
USER008	A5	A4	8
USER009	A3	A4	7
USER010	A7	A4	6
USER011	A1	A7	1
USER012	A10	A4	5
USER013	A1	A4	6
USER014	A9	A4	8
USER015	A2	A4	2
USER016	A10	A2	1
USER017	A3	A2	6
USER018	A5	A4	9
USER019	A3	A4	8
USER020	A9	A4	11
USER021	A6	A4	5

Table 3: Adjacency Matrix generated from processed CDR. Here users' return from work area to home area are not considered.

	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
A1	0	0	0	1	0	0	1	0	0	0
A2	0	0	0	2	0	0	1	0	0	0
A3	0	1	0	3	0	0	1	0	0	0
A4	0	0	0	0	0	0	0	0	0	0
A5	0	0	0	2	0	0	0	0	0	0
A6	0	0	0	1	0	0	0	0	0	0
A7	0	0	0	1	0	0	0	0	0	0
A8	0	1	0	0	0	0	0	0	0	0
A9	0	0	0	2	0	0	0	0	0	0
A10	0	0	0	2	0	0	0	0	0	0

As in India, tier-2 city sub urban areas are mostly sparsely populated mufassil areas, so very few numbers of mobile towers are placed in each such area.

2. CDR data are collected for at least 1 month for the selected area from a telecom provider with larger market share.

3. Assign each user (unique USER/IMEI) to its area code instead of cell tower. Now we can locate a user to an area code. It is an aggregation of users from cellId to *area_code*. It also allows to preserve privacy of individual users as after this step, no single user can be identified from the aggregated data.

4. We explore each week day's (Monday to Saturday) data separately. Remove all those users whose location did not change in the 24 hour period. So, we only have those users who goes out from his home location at least once in a week day.

5. Apply the proposed steps (discussed in next subsection) procedure from 7AM and up to 8PM for each day's CDR dataset.

5.2   Proposed Method

1. Generate the matrix as shown in table 2 by processing each CDR entries of each unique user with a granularity of 1 hour as detailed below:

   (a) Initially set user's home_area location as the current_area location as per CDR at 7 AM.
   $$home\_area := current\_area$$

   (b) Now check every 1 hour if user location has changed.

   (c) If yes then put the new location in *work_area* and *hour_stayed* as 1.

   (d) If in the following hour user's location remains same then only increase the *hour_stayed* by 1.

2. Remove all users with *hour_stayed* < 3 hour. Here we have used 3 hour as the threshold to select that location as his work_area location.

3. Count the number of users between each *home_area* and *work_area*.

4. Take average of the all above values over all the week days data in the dataset (20 days on a 4 week dataset)

5. Create adjacency matrix of size $n \times n$, where n is the number of sub areas, we have divided the whole region (In our example $n = 10$ as shown in table 3). Please note in the example adjacency matrix generated from table 2 we have not shown any user who are returning from work to home. In actual data there will be two entry for most of the users. One entry when they move from *home_area* to *work_area* and again in the afternoon when they return from work to home second entry gets generated. As we are only interested in creating un-directed network so, we sum up both count and generate upper triangular matrix for the purpose i.e. $a_{ij} = a_{ij} + a_{ji}$.

6. Draw a graph/network where vertices are the sub areas (a_i) and edges weights are the number of users between them i.e. $ew_{ij} = a_{ij} =$ number of users commuting between a_i and a_j on average per week day.

7. Add edge weights between two nodes (a_i and a_j) i.e. ew_{ij} and ew_{ji} to make the graph un-directional.

8. Assign node weights to the vertices as the sum of weights of all incident edges. $nw_{ai} = \sum_{a_j \in \Gamma_i} ew_{ij}$, where nw_{ai} is the node weight of the vertex a_i, Γ_i is the set of all edges connected to the node a_i.

9. Visualize the graph where edge thickness is proportional to the edge weights and vertex size is proportional to the node weight. The example graph is shown in figure3.

10. From the graph identify the large size vertices with thick edges from other vertices as probable edge-hubs/distribution centers and connected vertices are the end customer location areas that could be delivered using the proposed model.

As it is apparent from the generated graph in figure 3 that sub area "A4" should be considered for establishing edge-hub for distributing to most of the surrounding areas. In a real scenario, there will be millions of users but the number of sub-areas that we create shall always
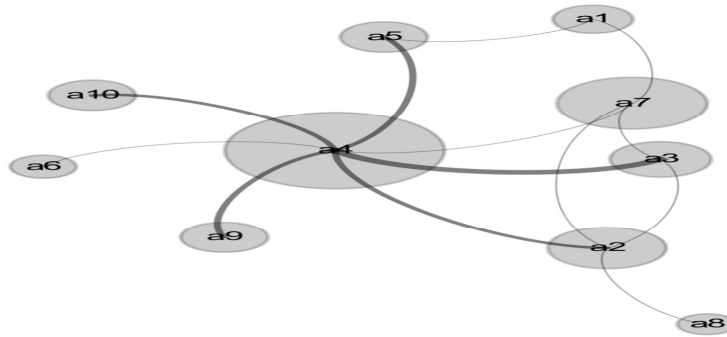
Figure 3: Network graph generated from the sample dataset to identify edge-hubs

be very small, maybe in 10s or 100s at max for the surrounding areas of some tier-2 or tier-3 city. So even then it would be possible to visually identify the edge-hubs to be considered for the proposed delivery model. Also as CDR are dynamic data and could be used in almost real-time so it is possible to observe the change in people's movement pattern if required.

6   Conclusion

A common phenomenon in most of the developing economies is the increase in smartphone use and as a result, a large number of online customers putting orders to the e-commerce companies. In developing countries like India, the most important problem, the e-commerce companies are facing with the double-digit growth every year is that of delivering the items to the remote parts of the country. As order volume and quantity both are less and geographically sparse in case of suburban and rural areas, home deliver the products become a costly affair. But developing countries are always very much price sensitive, so increase in delivery charges would not be an option as it might deter the customer from using e-commerce services at all. Upon close observation, it is seen that e-commerce delivery services are incurring higher cost during the 'last mile' to the greater suburban or muffasils of tier-2 and tier-3 cities. In this paper, a novel delivery model has been proposed that uses the local poor people who regularly travel to the tier-2 or tier-3 cities from surrounding remote areas for work. They have been utilized as a part-time delivery person for the orders that need to be delivered to their native areas for a little extra income. A method to identify the locations to place such delivery points where from nearby people collect items to deliver in their native places has also been proposed using telecom call detail record data. With proper implementation of the discussed model, e-commerce delivery could literally reach 'pan India' with an option to truly reach the huge untapped rural market.

References

[1] Qingqing Chen, Zheng Hu, Hang Su, Xiaosheng Tang, and Ke Yu. Understanding travel patterns of tourists from mobile phone data: A case study in hainan. In *Big Data and Smart Computing (BigComp), 2018 IEEE International Conference on*, pages 45–51. IEEE, 2018.

[2] Neeru Gupta and K Arora. Digital india: A roadmap for the development of rural india. *International Journal of Business Management*, 2(2):1333–1342, 2015.

[3] Mohanad Halaweh. Cash on delivery (cod) as an alternative payment method for e-commerce transactions: Analysis and implications. *International Journal of Sociotechnology and Knowledge Development (IJSKD)*, 10(4):1–12, 2018.

[4] Sibren Isaacman, Richard Becker, Ramón Cáceres, Stephen Kobourov, Margaret Martonosi, James Rowland, and Alexander Varshavsky. Identifying important places in people's lives from cellular network data. In *International Conference on Pervasive Computing*, pages 133–151. Springer, 2011.

[5] Abhinav Jha and Pinaki Roy. Rupay card. *Economic and Political Weekly*, 47(34):75–76, 2012.

[6] Nir Kshetri. Barriers to e-commerce and competitive business models in developing countries: A case study. *Electronic commerce research and applications*, 6(4):443–452, 2007.

[7] Dr Vinit Kumar. Pradhan mantri jan dhan yojana (pmjdy): Financial inclusion and inclusive growth in india. *International Journal of Scientific & Innovative Research Studies*, pages 2347–7660, 2015.

[8] Stanley Frederick WT Lim, Xin Jin, and Jagjit Singh Srai. Consumer-driven e-commerce: A literature review, design framework, and research agenda on last-mile logistics models. *International Journal of Physical Distribution & Logistics Management*, 48(3):308–332, 2018.

[9] Giridhar Maji, Lalmohan Dutta, and Soumya Sen. Targeted marketing and market share analysis on pos payment data using dw and olap. In *Emerging Technologies in Data Mining and Information Security*, pages 189–199. Springer, 2019.

[10] Giridhar Maji and Soumya Sen. A data warehouse based analysis on cdr to depict market share of different mobile brands. In *India Conference (INDICON), 2015 Annual IEEE*, pages 1–6. IEEE, 2015.

[11] Giridhar Maji and Soumya Sen. Data warehouse based analysis on cdr to retain and acquire customers by targeted marketing. In *Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), 2016 5th International Conference on*, pages 221–227. IEEE, 2016.

[12] Sharmistha Mandal and Giridhar Maji. Integrating telecom cdr and customer data from different operational databases and data warehouses into a central data warehouse for business analysis. *Int. Journal of Engineering Research & Technology*, 5(2):516–523, 2016.

[13] S Muthukumar and N Muthu. The indian kaleidoscope: emerging trends in m-commerce. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(1):50–56, 2015.

[14] Mikko Punakivi, Hannu Yrjölä, and Jan Holmström. Solving the last mile issue: reception box or delivery box? *International Journal of Physical Distribution & Logistics Management*, 31(6):427–439, 2001.

[15] STATISTICA. Retail e-commerce sales in india from 2016 to 2022 (in million u.s. dollars). *Available at* *https: // www. statista. com/ statistics/ 289770/ india-retail-e-commerce-sales/* , April 2018.

[16] Hannu Yrjola. Physical distribution considerations for electronic grocery shopping. *International Journal of Physical Distribution & Logistics Management*, 31(10):746–761, 2001.