Development of origin–destination matrices using mobile phone call data

Development of Origin-Destination Matrices Using Mobile Phone Call Data: A Simulation Based Approach

Md. Shahadat Iqbal
Department of Civil Engineering.
Bangladesh University of Engineering and Technology, Dhaka 1000, Bangladesh
Shahadat.buet05@gmail.com

Charisma F. Choudhury*
Institute for Transport Studies
University of Leeds, Leeds LS2 9BJ, UK
cfc@alum.mit.edu

Pu Wang
School of Traffic and Transportation Engineering
Central South University, Hunan 410000, P.R. China

Marta C. Gonza´lez
Department of Civil and Environmental Engineering,
Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Word Count	Tables and Figures	13 x 250 = 3250
	Word Count	3814
	Total	7064

*Corresponding Author

Abstract

In this research, we propose a methodology to develop OD matrices using mobile phone Call Detail Records (CDR), which consist of time stamped tower locations with caller IDs, and limited traffic counts. CDR from 2.87 million users from Dhaka, Bangladesh over a month and traffic counts from 13 key locations of the city over 3 days of the same period are used in this regard. The individual movement patterns within certain time windows are extracted first from CDR to generate tower-to-tower *transient* OD matrices. These are then associated with corresponding nodes of the traffic network and used as seed-OD matrices in a microscopic traffic simulator. An optimization based approach, which aims to minimize the differences between observed and simulated traffic counts at selected locations, is deployed to determine scaling factors and the actual OD matrix is derived. The applicability of the methodology is supported by a validation study.

Keywords: Mobile phone, Origin-Destination, Video Count, Traffic Microsimulation

1. Background

Reliable Origin-Destination (OD) matrices are critical inputs for analyzing transportation initiatives. Traditional approaches of developing OD matrices rely on roadside and household surveys, and/or traffic counts. The roadside and household surveys for origin destination involve expensive data collection and thereby have limited sample sizes and lower update frequencies. Moreover, they are prone to sampling biases and reporting errors (e.g.1,2,3). Estimation of reliable OD matrices from traffic link count data on the other hand is extremely challenging since very often the data is limited in extent and can lead to multiple plausible non-unique OD matrices (4,5). A number of Bayesian methods (e.g.6,7,8), Generalized Least Squares approaches (e.g.9,*10*), Maximum Likelihood Approaches (*11*), and Correlation Methods (e.g.*12,13,14*) have been used to tackle the indeterminacy problem. These approaches typically use *target* matrices based on prior information for generating the plausible route flows and are very sensitive to this prior information as well as to the chosen methodology (*15*). More recent approaches for OD estimation include automated registration plate scanners (*16*) and mobile traffic sensors such as portable GPS devices (e.g.*17,18,19*) . The practical successes of these approaches have however been limited due to high installation costs of the license plate readers and the low penetration rates of GPS devices (especially in developing countries).

Mobile phone users on the other hand also leave footprints of their approximate locations whenever they make a call or send an SMS. Over the last decade, mobile phone penetration rates have increased manifold both in developed and developing countries: the current penetration rates being 128% and 89% in developed and developing countries respectively (*20*). Subsequently, mobile phone data has emerged as a very promising source of data for transportation researchers. In recent years, mobile phone data have been used for human travel pattern visualization (e.g. *21,22,23*), mobility pattern extraction (e.g. *24,25,26,27,28,29*), route choice modeling (e.g. *30,31*), traffic model calibration (e.g. *32*), traffic flow estimation (*33*) to name a few. There have been several limited scale researches to explore the feasibility of application of mobile phone data for OD estimation as well. Wang et al. (*34*) for instance use a correlation based approach to dynamically update a prior OD matrix using time difference of phone signal receipt times of base stations and Caceras et al. (*35*) use a GSM network simulator to simulate the detailed movements of phones that are turned on. But both of these feasibility studies are based on synthetic data in small networks and the practical application is challenging given the need to collect and process detailed location data (which are currently processed by the mobile phone companies for load management purposes but are not stored). The potential estimate OD matrices using mobile phone Call Detail Records (CDR) (which are stored by operators for billing purposes and hence more readily available) have also been explored (e.g. *36,37,38*). Mellegård et al. (*36*) have developed an algorithm to assign mobile phone towers extracted from CDR to traffic nodes and Calabrese et al. (*37*) have proposed a methodology to reduce the noise in the CDR data but both studies have focused more on computation issues and the relationship between the mobile phone OD and the traffic OD have not been explored in detail. Wang et al. (*38*) have used an analytical model to scale up the ODs derived from CDR by using the population, mode choice probabilities and vehicle occupancy and usage ratios and have validated it using probe vehicle data. The methodology however relies heavily on availability of traffic and demographic data in high spatial resolution which may not be always available, particularly in developing countries.

In this research, we propose a methodology to develop OD matrices using mobile phone CDR and limited traffic counts. CDR from 2.87 million users from Dhaka, Bangladesh over a month are used to generate the OD patterns on different time periods and traffic counts from 13 key locations of the city over a limited time are used to scale it up to derive the actual ODs using a microscopic traffic simulator. The methodology is particularly useful in situations when there is limited availability of high resolution traffic and demographic data. The ODs are validated by comparing the simulated and observed traffic counts of a different location (which has not been used for calibration).

The rest of the paper is organized as follows. First we describe the data followed by the methodology used for development of the OD matrix. The estimation and validation results are presented next. We conclude with the summary of findings and directions for future research.

2. Data

2.1 Study Area

The central part of the Dhaka city has been selected as the study area and the major roads in the network has been coded. This consists of 67 nodes and 215 links covering an area of about 300km^2 with a population of about 10.7million (*39*). The average trip production rate is 2.74 per person per day with significant portions of walking (19.8%) and non-motorized transport trips (38.3%) (*39*).The traffic is subjected to severe congestion in most parts of the day, the average speed being only 17km/hr[1].

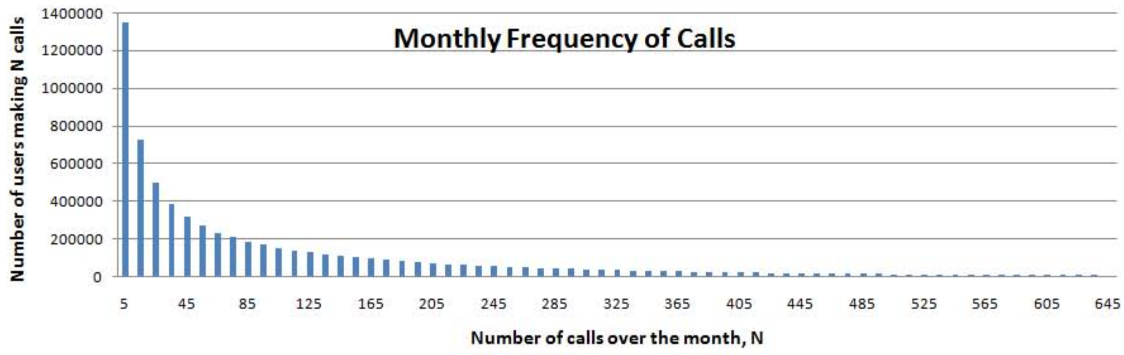
The mobile phone penetration rate is approximated to be more than 90% in Dhaka (66.36% being the national average) and Grameenphone Ltd. has the highest market share with 42.7m mobile phone subscribers nationwide (*40*).

2.2 CDR Data

The CDR data, collected from Grameenphone Ltd, consists of calls from 6.9 million users (which are more than 65% of the population of the study area) over a month. This comprises of 971.33 million anonymized call records in total made in between June 19, 2012 and July 18, 2012. The majority of the users (63%) have made 100 calls or less over the month. The frequencies of users making certain number of calls over the month and on a randomly selected

[1] Excluding the non-motorized vehicles which are restricted from entering the major roads

day (15th July, 2012) are presented in Figure 1. It may be noted that no demographic data related to the phone users are available.



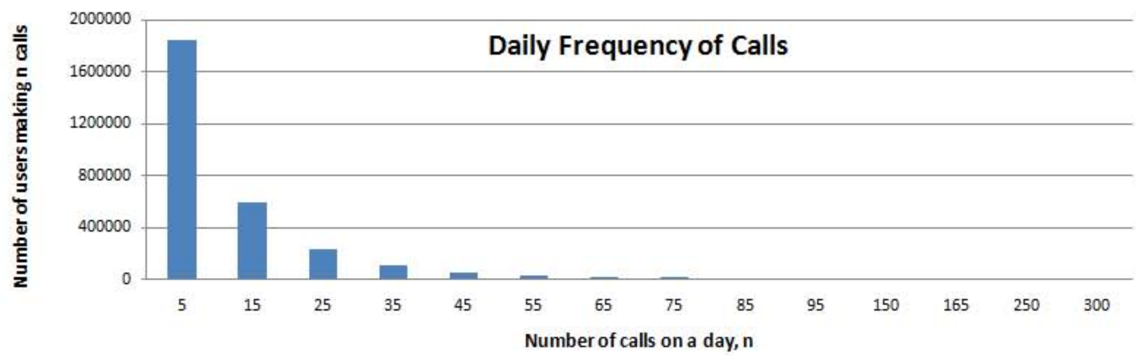


Figure 1: Frequency of calls per user

2.3 Traffic Count Data

Video data, collected from 13 key locations of Dhaka city network over 3 days (12th, 15th and 17th July 2012) have been used in this study to extract the traffic counts[2]. The locations (shown in Figure 2) have been selected such that they cover the major roads (links) of Dhaka city with flows from major generators and governed by the availability of foot over bridges for mounting video cameras. Since MITSIMLab is developed for lane-based motorized traffic, care has been taken to avoid roads that have high percentages of non-motorized transport and where lane-discipline is not strictly followed. The data has been collected for 8 hrs (8.00 am to 12.00 noon and 3.00 pm to 7.00pm) and analyzed using the software TRAZER (*41*) to generate classified vehicle counts. Due to inclement weather and poor visibility some portion of the data is non-usable though. Moreover, TRAZER (which is the only commercial software that can deal with mixed traffic streams with *'weak'* lane discipline) has high misspecification rates in presence of high congestion levels and in those cases, manual counting has been performed instead.

[2] There are no loop detectors or any other automatic traffic counters in Dhaka

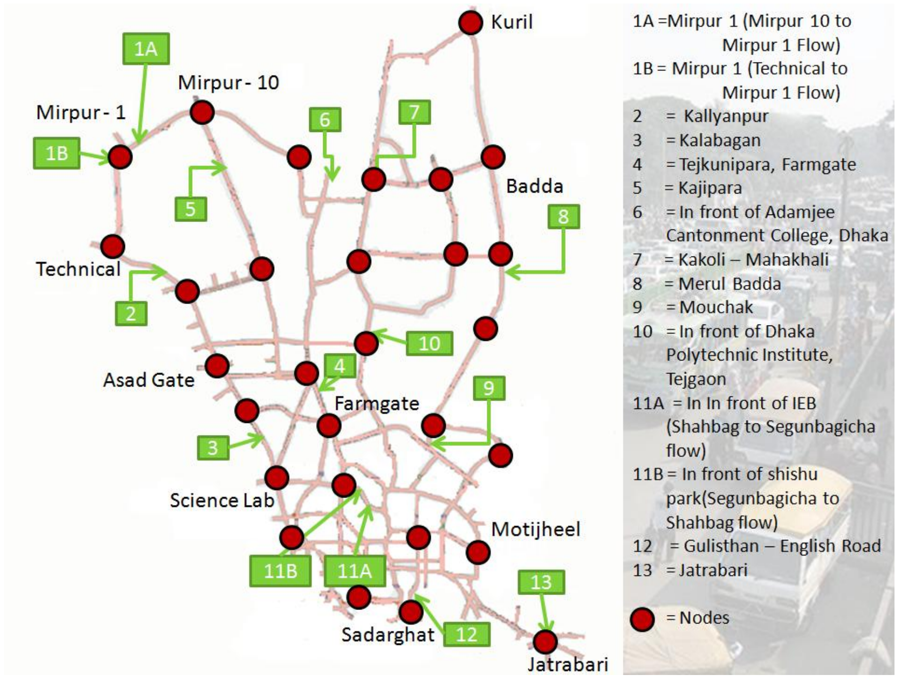


Figure 2: Locations of video data collection and position of OD generating nodes

3. Methodology

Each entry in the CDR contains unique caller id (anonymized), the date and time of the call, call duration and latitude and longitude of the Base Transceiver Station (BTS). A snapshot of the data is presented in Figure 1. As seen in the figure, if a person traverses within the city boundary and uses his/her phone from different locations that is captured in the CDR. CDR can thus provide an abstraction of his/her physical displacements over time (Figure 3).

ID	Call Date	Call Time	Duration	Latitude	Longitude
AH03JAC8AAAbXtAId	20120701	09:34:19	18	23.8153	90.4181
AAH03JABiAAJKnPAa5	20120707	06:15:20	109	23.8139	90.3986
AAH03JABiAAJKnPAa5	20120707	09:03:06	109	23.7042	90.4297
AAH03JABiAAJKnPAa5	20120707	10:34:19	16	23.6989	90.4353
AAH03JABiAAJKnPAa5	20120707	18:44:53	154	23.6989	90.4353
AAH03JABiAAJKnPAa5	20120707	20:00:08	154	23.8092	90.4089
AAH03JAC5AAAdAYAE	20120701	09:15:05	62	23.7428	90.4164
AAH03JAC+AAAcVKAC	20120707	08:56:34	242	23.7908	90.3753
AAH03JAC+AAAcVKAC	20120701	18:03:06	36	23.9300	90.2794
AAH03JAC5AAAdAYAA	20120701	11:15:55	12	23.7428	90.4164

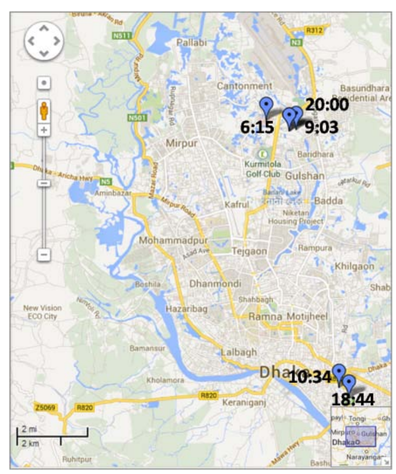


Figure 3: An excerpt from CDR data (entries of the same user are highlighted) and locations of a random user “AAH03JABiAAJKnPAa5” throughout the day as observed in data

However, in the CDR data, a user's location information is lost when he/she does not use his/her phone. As shown in Figure 4, according to the CDR, a user may be observed to move from zone B to zone C, but his/her initial origin (O) and final destination (D) may actually be located in zone A and zone D. In such cases, a segment of the trip information is unobserved in the CDR. However, the mobile phone call records enable us to capture the *transient* origins and destinations which still retain a large portion of the actual ODs. Thus, we use the concept of transient origin destination (*t*-OD) matrix (as used by Wang et al. (*38*)), which uses the mobile phone data to efficiently and economically capture the pattern of travel demand.

Figure 4: Actual vs. Transient OD

The second source of data used in this research is classified traffic counts extracted from video recordings collected from 13 key locations of Dhaka. These counts represent the *ground truth* but are more expensive to collect[3] and limited in extent (only 3 days). This limited point source data therefore cannot be used as a stand-alone source to reliably capture the OD pattern.

In this research, we therefore plan to combine the two data sources. The OD pattern is generated using the CDR data and scaled up to match the traffic counts. The scaling factors are determined using a microscopic traffic simulator platform MITSIMLab (*42*) using an optimization based approach which aims to minimize the differences between observed and simulated traffic counts at the points where the traffic counts are available.

The methodology is summarized in Figure 5 and described in the subsequent sections.

[3] There are no detectors or any other traffic count mechanisms in Dhaka

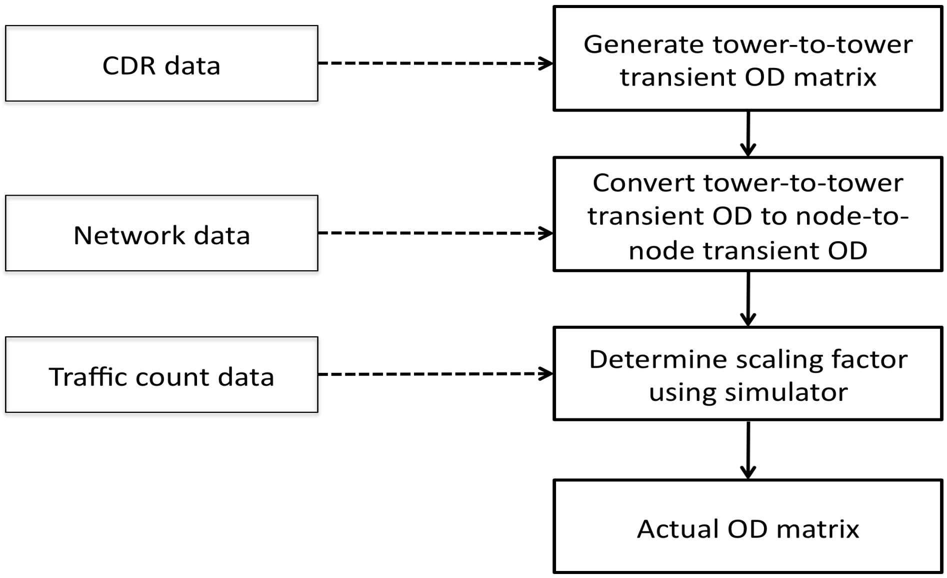


Figure 5: Framework for developing OD Matrix

3.1 Generation of tower-to-tower transient OD matrix

The time-stamped BTS tower locations of each user are first extracted from the mobile phone CDR data and used for generating tower-to-tower transient OD matrix. The CDR however only contains sparse and irregular records (*28*), in which user displacements (consecutive non-identical locations) are usually observed with long travel intervals i.e. the first location may be observed at 8:56 and next location may be observed at 18:03 with no information about intermediate locations (if any) or the time when the trip in between these two locations have been made.

Another limitation is the CDR data often records changes in towers in spite of no actual displacement (as the operator balances call traffic among adjacent towers). To better identify timing and origin-destinations of specific trips and reduce the number of *false displacements*, we therefore extract displacements that have occurred within a specific *time window*. A lower bound in the time window (10 minutes) is imposed to reduce the number of *false displacements* without affecting the number of physical displacements occurring within short intervals. An upper bound in the time window (1 hr) is imposed to ensure that meaningful numbers of trips are retained. Therefore, a person trip is recorded if in the CDR, subsequent entries of the same user indicate a displacement (change in tower) with a time difference of more than 10 minutes but less than 1 hour.

Further, both call volumes (from CDR data) and traffic volumes (from traffic counts) had significant variations throughout the day. Based on correlation analysis of total mobile call volumes and total traffic counts (Figure 6), four time periods (7:00-9:00, 9:00-12:00, 15:00-17:00 and 17:00-19:00), have been chosen for analysis.

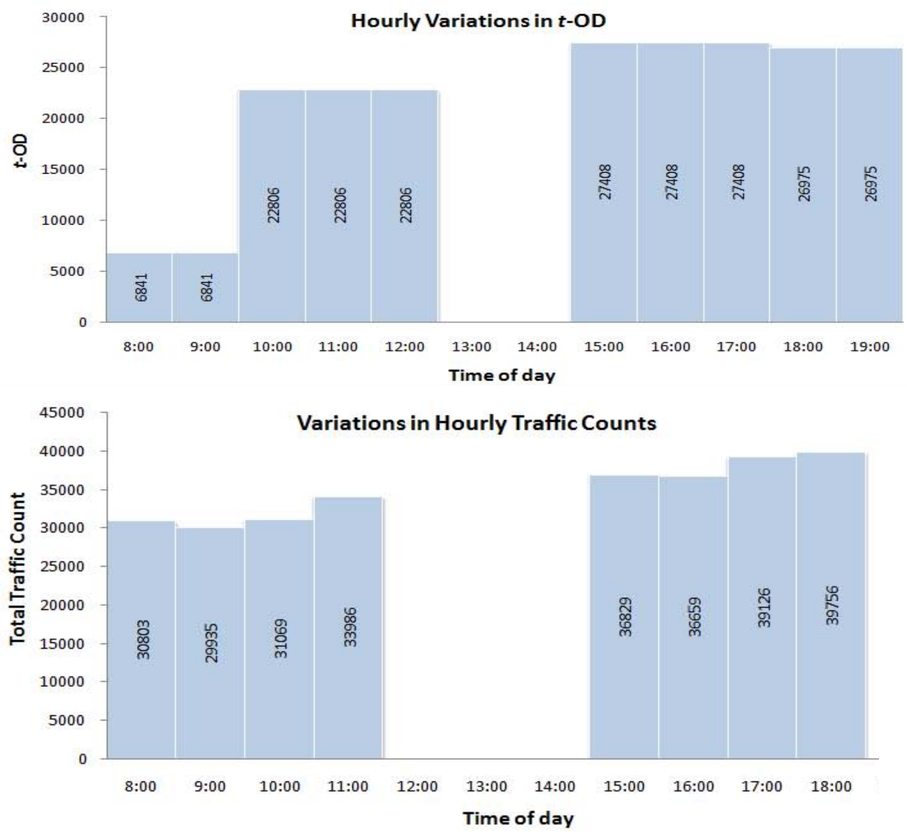


Figure 6: Hourly variations a. traffic count b. transient ODs from mobile call records

3.2 Conversion of tower-to-tower t-OD to node-to-node t-OD

For application of the *t*-ODs in traffic analyses, the origin and destination towers need to be associated with corresponding nodes of the traffic network. The typical tower coverage area can be represented as a combination of three hyperbolas (Figure 7), the size varying depending on tower height, terrain, locations of adjacent towers and number of users active in the proximity (which can vary dynamically).

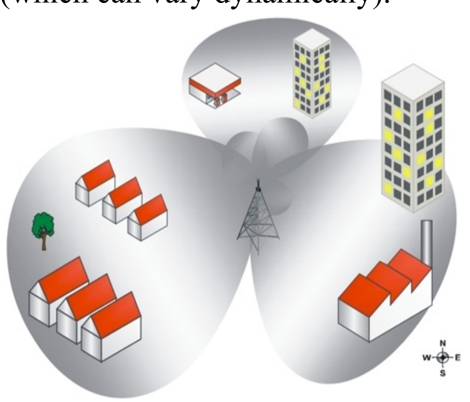


Figure 7: Typical coverage area of a tower (http://www.truteq.co.za/tips_gsm/)

The population density in the chosen study area is very high (more than 8111 inhabitants/sq. km (*44*) and the tower locations are very close to each other (1 km on average). Because of the high user density, it can be assumed that the area between two towers is equally split among the two towers (Figure 8) that is, each tower t has a coverage area (A_t) approximately defined by a circle of radius $0.5l$, where l is the tower-to-tower distance.

Tower 6 and Node 3 need to be added to Figure

ID	Call Date	Call Time	Origin Tower	Destination Tower
AAH03JA	20120718	15:54	6	1
AAH03JA	20120718	16:13	1	2
AAH03JA	20120718	16:15	2	1
AAH03JA	20120718	18:53	1	6
AAH03JA	20120718	20:49	6	1
AAH03JA	20120718	23:41	1	6

a. Tower-to-tower OD

ID	Call Time	Origin		Destination	
		Tower	Candidate Node	Tower	Candidate Node
AAH03JA	14:54	6	3	1	1
AAH03JA	16:13	1	1	2	2 Or 1
AAH03JA	16:15	2	2 Or 1	1	1
AAH03JA	18:53	1	1	6	3
AAH03JA	20:49	6	3	1	1
AAH03JA	23:41	1	1	6	3

b. Intermediate OD with candidate nodes

ID	Call Time	Origin Node	Destination Node
AAH03JA	14:54	3	1
AAH03JA	16:13	1	1
AAH03JA	16:15	1	1
AAH03JA	18:53	1	3
AAH03JA	20:49	3	1
AAH03JA	23:41	1	3

c. Node-to-node OD

Figure 8: Example of tower to node allocation

If a unique traffic node i overlaps with A_t, the calls handled by t are associated with node i (as in the case of Tower 1in Figure 6). However, if A_t has two (or more) candidate nodes for association, then the candidate nodes are ranked based on the proportion of A_t feeding to each node. That is, the node serving greatest portion of A_t is ranked 1, the node serving second highest portion of A_t is ranked 2, etc. For example, in Figure 6, network connectivity (feeder roads) and topography (presence of a canal with no crossing facility in the vicinity) denote that Node 1 and Node 2 are candidate nodes for association with Tower 2. As the major portion of A_t is connected to Node 2 and the remaining portion is connected to Node 1, they are ranked 1 and 2 respectively for Tower 2. The data format after this step is presented in Figure 7b. As seen in the figure, this typically consists of call records associated with unique nodes and some calls associated with *multiple candidate nodes*. The calls are then sorted and ranked based on the frequency of the unique nodes used by each user. The frequency of occurrence of the candidate nodes are

compared and used as the basis of replacement. For example, frequency analysis of User "AAH03JA" indicates a higher frequency of Node 1. Therefore, in cases where there are ambiguities between Nodes 2 and 1, Node 1 is used (for this particular user).
The same process is used for all users and node-to-node *t*-OD matrices for each time period of each day are derived.

3.3 Finding the scaling factor and determining the actual OD matrix
As discussed, the node-to-node *t*-OD matrix ($t\text{-}OD_{ij}$) provides the trip patterns for developing the actual OD matrix (OD_{ij}). However, in order to determine the actual OD matrix, the *t*-OD needs to be scaled to match the real traffic flows. A scaling factor β_{ij} is used in this regard:

$$OD_{ij} = \sum_{ij} (t\text{-}OD_{ij}) * \beta_{ij}$$

It may be noted that β_{ij} takes into account the market penetration rates (i.e. not every user has a mobile phone or uses the specific service provider), the mobile phone non-usage issue (i.e. mobile phone calls are not made from every location traversed by the user), the vehicle usage issue (i.e. users may not use cars for every trip). The potential error introduced due to *false displacement* (described in Section 2.1) is also accounted for in the scaling factors.
The scaling factors are determined using the open-sourced microscopic traffic simulator platform MITSIMLab (*42*) by applying an optimization based approach. The movements of vehicles in MITSIMLab are dictated by driving behavior models based on decision theories and estimated with detailed trajectory data using econometric approaches. Route choices of drivers are based on a discrete choice based probabilistic model where the utilities of selecting and re-evaluating routes are functions of path attributes, such as path travel times and freeway bias (see *43* for details). The inputs of the simulator include network data, driving behavior parameters and OD matrix. The generated outputs include traffic flow at specified locations in the network.

The node-to-node OD matrix derived from the mobile phone data are provided as the initial or seed-OD in this case. The simulated traffic flows are compared with the actual traffic flows extracted from video recordings. The objective function seeks to minimize the difference between the actual and simulated traffic flows in each location by changing the scaling factors. The optimization problem can be represented as follows:

$$minimize, Z = \sum_{k=1}^{K} (V_{actual}^{k} - V_{simulated}^{k})^2 \qquad (1)$$

$$\text{Such that, } OD_{ij,t} = \sum_{i,j=1}^{N} t\text{-}OD_{ij,t} * \beta_{ij,t}$$

Where,
$V_{simulated}^{k}$ = Traffic flow of link *k* of the road network from simulation
$OD_{ij,t}$ = Actual OD between nodes *i* and *j* in time period *t*
$t\text{-}OD_{ij,t}$ = Transient OD between nodes *i* and *j* in time period *t*
$\beta_{ij,t}$ = Scaling factor associated with the node pair *i* and *j* and time period *t*

K = Total number of links for which traffic flow data is available
N = Total number of nodes in the network

However, to make the optimization problem more tractable, group-wise scaling factors are used rather than an individual scaling factor for each OD pair. The grouping is based on the analyses of the CDR data. This simplifies the problem as follows:

$$minimize, Z = \sum_{k=1}^{K} (V_{actual}^{k} - V_{simulated}^{k})^2 \quad (2)$$

$$\text{Such that, } OD_{ij,t} = \sum_{m=1}^{M} t\text{-}OD_{ij,t}^{m} * \beta_t^m$$

Where,

$t\text{-}OD_{ij,t}^{m}$ = Transient OD between node pair *i* and *j* in time period *t* where the node pair *i,j* belong to group *m*
β_t^m = Scaling factor for group m and time period *t*
M = Total number of groups of OD-pairs

4. Results

The mobile phone network within the study area comprises of 1360 towers which have been assigned to 29 OD generating nodes (812 OD pairs). Out of the one month CDR data, the weekend data have been discarded. For each day, the calls of each user originating from two different towers in each of the time period have been extracted. After application of the transient trip definitions (displacements occurring more than 10mins but less than 1hr apart) and the tower to node conversion rules (elaborated in Section 3.2), the node-to-node *t*-ODs are derived. The total number of node-to-node *t*-ODs are presented in Table 1.

Table 1: Node-to-node *t*-OD

Time Period	Time	***t*-OD**	
		Total Over the Month[4]	**Weekday Average**
1	7:00-9:00	397355	13681.86
2	9:00-12:00	1915417	68418.48
3	15:00-17:00	2255859	82226.05
4	17:00-19:00	1549109	53950.57

[4] Includes weekends

Analyses of the node-to-node transient flows indicate that the flows between adjacent nodes are substantially higher than those between non-adjacent nodes (Figure 9). This is reasonable since given the low travel speed in Dhaka, a traveler may not be able to move very far in the 50min time window and the *t*-ODs mostly capture segments of a longer trip. However, part of it may also be due to the *false displacement* problem discussed in section 3.1. Therefore, the OD-pairs have been divided into two groups (adjacent and non-adjacent nodes) and the objective function to determine scaling factors has been formulated as follows:

$$minimize, Z = \sum_{k=1}^{K}(V_{actual}^{k} - V_{simulated}^{k})^2 \qquad (3)$$

Such that, $OD_{ij,t} = \sum_{adj} t\text{-}OD_{ij,t}^{adj} * \beta_t^{adj} + \sum_{non\text{-}adj} t\text{-}OD_{ij,t}^{non\text{-}adj} * \beta_t^{non\text{-}adj}$

Where,

$t\text{-}OD_{ij}^{adj}$ = Transient OD between node pair *i* and *j* in time period *t* where the node pair *i,j* are adjacent nodes

$t\text{-}OD_{ij}^{non\text{-}adj}$ = Transient OD between node pair *i* and *j* in time period *t* where the node pair *i,j* are non-adjacent nodes

$\beta_t^{adj}, \beta_t^{non\text{-}adj}$ = Scaling factors for time period *t* and adjacent and non-adjacent nodes respectively

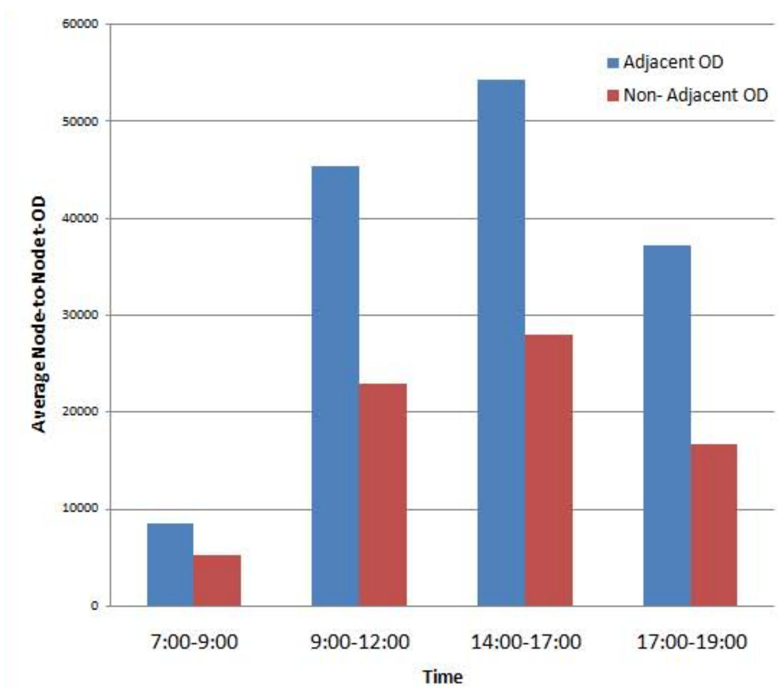


Figure 9: Comparison of *t*-ODs between adjacent and non-adjacent nodes

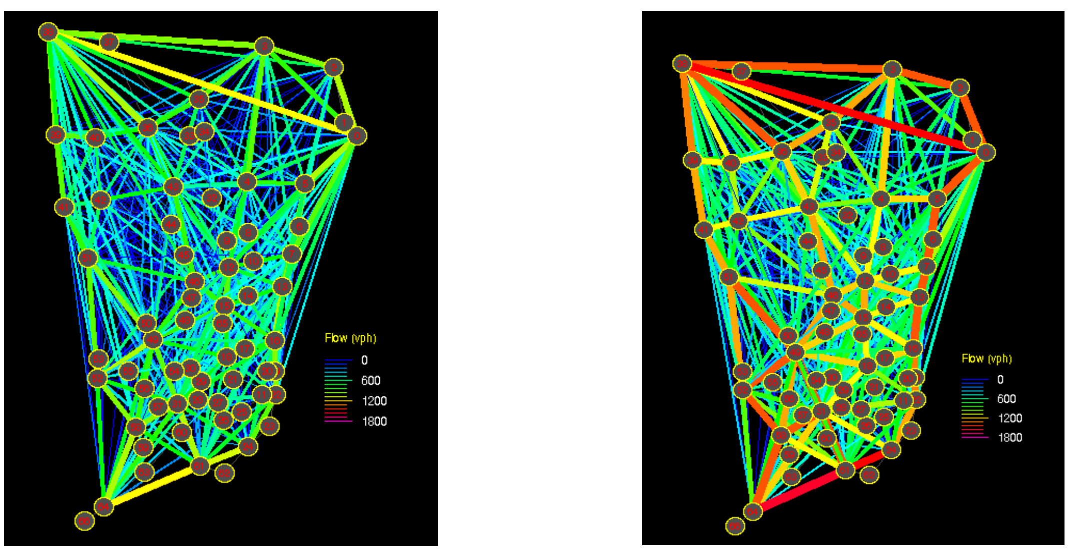
This yielded eight scaling factors in total that needed to be estimated from the simulation runs of MITSIMLab. Running the optimization process in MATLAB (that invokes MITSIMLab) and using a BOX algorithm (*45*), the following values of scaling factors have been derived.

Table 2: Scaling Factors

Time Period	OD Type	Scaling Factor
7:00-9:00	Adjacent	6.787
	Non-adjacent	1.712
9:00-12:00	Adjacent	0.971
	Non-adjacent	0.345
15:00-17:00	Adjacent	1.647
	Non-adjacent	3.407
17:00-19:00	Adjacent	9.404
	Non-adjacent	6.779

It is interesting to note that the scaling factors for adjacent nodes are higher than those of non-adjacent in all time periods other than 15:00-17:00. This does not however indicate that most of the actual trips are to the adjacent nodes (since a full trip may consist of several segments each represented by a separate *t*-OD).

The graphical representation of the *t*-ODs and actual ODs across the network for one of the time periods and the variations for an example node are presented in Figures 10 and 11 respectively.



a. t-OD b. actual OD

Figure 10: *t*-ODs and actual ODs across the network for 7:00-9:00

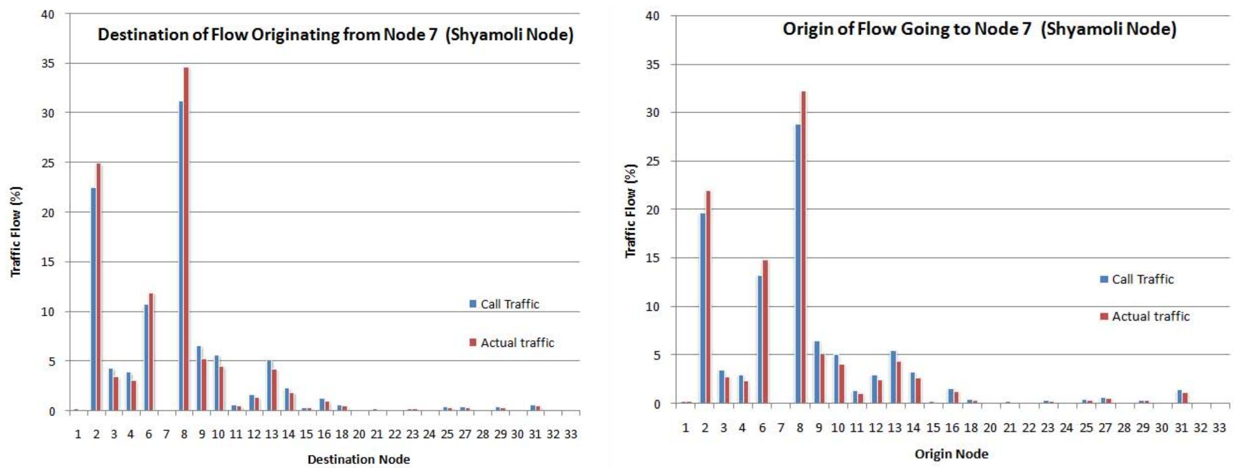


Figure 11: Example of Transient and Actual Traffic Flows To and From a Node (Shyamoli) between 7:00-9:00.

5. Validation

In addition to the aggregate data used for calibration, traffic counts are collected from four additional locations on a different day. For validation purposes, the scaled up ODs have been applied to simulate the traffic between 9:00-12:00 in MITSIMLab and the simulated traffic counts are compared against the observed counts from these locations. In order to quantify the prediction error, Root Mean Square Error and Root Mean Square Percent Errors have been calculated and are found to be 335.09 and 13.59% respectively.

6. Conclusion

The main outcome of this research is the methodology for development of the OD matrix using mobile phone CDR and limited traffic count data. The strengths of both data sources are utilized in this approach: the trip patterns are extracted from mobile phones and the ground truth traffic scenario are derived from the counts. The methodology is demonstrated using data collected from Dhaka.

There are several limitations of the current research though. Firstly, in this research a simplified objective function with grouped scaling factors has been used. This overlooks the heterogeneity in call rates from different locations (e.g., more calls may be generated to and from railway stations compared to and from offices with land telephone lines, etc.). A more detailed classification of scaling factor can be used to overcome this bias and may yield better results. Moreover, in this particular context, detailed network data and extensive calibration data were not available which may have increased the simulation errors and affected the validation results. However, initial validation results indicate promising success in real life application by transport planners and managers.

Since CDR is already recorded by mobile phone companies for billing purposes, the approach is more economic than the traditional approaches which rely on expensive household surveys and/or extensive traffic counts. It is also convenient for periodic update of the OD matrix and extendable for dynamic OD estimation. This method is particularly effective for generating complex OD matrix where land use pattern is heterogeneous and asymmetry in travelling pattern prevails throughout the day but there is a limitation of traditional data sources.

Acknowledgment

The data provided for the research has been provided by Grameenphone Ltd. , Bangladesh. The funding for this research was provided by Faculty for the Future Program of Schlumberger Foundation and Higher Education Enhancement Project of the University Grants Commission of Bangladesh and the World Bank.

References

1. Hajek, J. J. (1977). *Optimal sample size of roadside-interview origin-destination surveys* (No. RR 208).
2. Kuwahara, M., and Sullivan, E. C. (1987). Estimating origin-destination matrices from roadside survey data. *Transportation Research Part B*,*21*(3), 233-248.
3. Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. Public Opinion Quarterly, 70(5), 646-675.
4. Lo, H. P., Zhang, N., and Lam, W. H. (1996). Estimation of an origin-destination matrix with random link choice proportions: a statistical approach.*Transportation Research Part B*, *30*(4), 309-324.
5. Van Zuylen, H. J., and Willumsen, L. G. (1980). The most likely trip matrix estimated from traffic counts. *Transportation Research Part B*,*14*(3), 281-293.
6. Maher, M. (1983). Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research Part B, 20 (6)*, 435–447.
7. Tebaldi, C., West, M. (1998). Bayesian inference on network traffic using link count data (with discussion). *Journal of the American Statistical Association,93,* 557–576.
8. Li, B. (2005). Bayesian inference for origin–destination matrices of transport networks using the EM algorithm. *Technometrics 47 (4)*, 399–408.
9. Cascetta, E. (1984). Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B, 18(4–5)*, 289–299.
10. Bell, M. (1991). The estimation of origin–destination matrices by constrained generalized least squares. *Transportation Research Part B, 25 (1)*, 13–22.
11. Spiess, H. (1987). A maximum likelihood model for estimating origin-destination matrices, *Transportation Research Part B,21(5)*, 395-412.
12. Vardi, Y. (1996). Network tomography: estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association,91,* 365–377.
13. Hazelton, M.L. (2000). Estimation of Origin–Destination matrices from link flows on uncongested networks. *Transportation Research Part B, 34 (7)*, 549–566.

14. Hazelton, M.L. (2003). Some comments on origin–destination matrix estimation. *Transportation Research Part A, 37 (10)*, 811–822.
15. Hazelton, M.L., 2001b. Inference for origin–destination matrices: estimation, reconstruction and prediction. *Transportation Research Part B, 35 (7),* 667–676.
16. Castillo, E., Menéndez, J., Jiménez, P. (2008). Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transportation Research Part B* ,42 (5), 455–481.
17. Parry, K., & Hazelton, M. L. (2012). Estimation of origin–destination matrices from link counts and sporadic routing data. *Transportation Research Part B, 46(1),* 175-188.
18. Morimura, T., and Kato, S. (2012). Statistical origin-destination generation with multiple sources. 21st International Conference on In Pattern Recognition (ICPR), November 11-15, 2012. Tsukuba, Japan.
19. Herrera, J., Work D. B., Herring R., Ban X., Jacobson Q., Bayen A. (2010). Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment, *Transportation Research Part C: Emerging Technologies, 18(4)*, 568-583.
20. http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013.pdf [accessed 20July, 2013]
21. Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., and Ratti, C. (2010). Activity - Aware Map: Identifying human daily activity pattern using mobile phone data, *Human Behavior Understanding, 6219(3),* 14-25,Springer Berlin / Heidelberg.
22. Phithakkitnukoon, S., and Ratti, C., (2011), Inferring Asymmetry of Inhabitant Flow using Call Detail Records, *Journal of Advances in Information Technology*, *2 (4)*, 239-249.
23. Reades, J., Calabrese, F., and Ratti, C. (2009). Eigenplaces: analyzing cities using the space-time structure of the mobile phone network, *Environment and Planning B: Planning and Design, 36(5),* pp. 824-836.
24. Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., and González, M. C. (2012). Understanding Road Usage Patterns in Urban Areas. *Scientific reports, 2*.
25. G onzález, M. C., Hidalgo, C. A., and Barabási, A. L.(2008).Understanding individual human mobility patterns, *Nature, 453,* 779–782.
26. Song, C, Koren, T, Wang, P, and Barabási, A. L. (2010). Modelling the scaling properties of human mobility, *Nature Physics, 6*, 818–823.
27. Simini, F., Gonza´lez, M. C., Maritan, A., and Baraba´si, A. L.(2012). A universal model for mobility and migration patterns, *Nature, 484*, 96–100.
28. Candia, J., González, M. C., Wang, P., Schoenharl, T., Madey, G., and Barabási, A. L. (2008). Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical, 41(22)*, 224015.
29. Sevtsuk, A., and Ratti, C. (2010). Does Urban Mobility Have a Daily Routine? Learning from Aggregate Data of Mobile Networks, *Journal of Urban Technology, 17 (1)*, 41-60.
30. Schlaich, J., Otterstätter, T., Friedrich, M., 2010, Generating Trajectories from Mobile Phone Data, TRB 89th Annual Meeting Compendium of Papers, Transportation Research Board of the National Academies, Washington, D.C., USA.
31. Becker, R.A., Caceres, R., Hanson, K., Loh, J.M., Urbanek, S., Varshavsky, A., Volinsky, C., Ave, P., Park, F., 2011. Route classification using cellular handoff patterns. In: Proceedings of the 13th International Conference on Ubiquitous Computing. ACM, Beijing, China.

32. Bolla, R., Davoli, F., and Giordano, A. (2000). Estimating road traffic parameters from mobile communications. In *Proceedings 7th World Congress on ITS, Turin, Italy*.
33. Demissie, M. G., de Almeida Correia, G. H., and Bento, C. (2013). Intelligent road traffic status detection system through cellular networks handover information: An exploratory study. Transportation *Research Part C: Emerging Technologies, 32*, 76-88.
34. Wang J., Wang D. Song X. Sun Di. (2011). Dynamic OD Expansion Method Based on Mobile Phone Location, Fourth International Conference on Intelligent Computation Technology and Automation, Shenzhen, China.
35. Caceres, N., Wideberg, J. P., and Benitez, F. G. (2007). Deriving origin destination data from a mobile phone network. *Intelligent Transport Systems, IET*, *1*(1), 15-26.
36. Mellegard, E., Moritz, S., and Zahoor, M. (2011, December). Origin/Destination-estimation using cellular network data. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on* (pp. 891-896). IEEE.
37. Calabrese F., Lorenzo G. D., Liu L. and Ratti C. (2011). Estimating Origin-Destination Flows using Mobile phone Location Data. IEEE Pervasive Computing, vol. XX, no. XX, 200XX, pp. 36–43.
38. Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., and González, M. C. (2012). Understanding Road Usage Patterns in Urban Areas. *Scientific reports*, *2*.
39. DHUTS. (2010). Dhaka Urban Transport Network Development Study, Draft Final Report. Prepared by Katahira and Engineers International, Oriental Consultants Co. Ltd., and Mitsubishi Research Institute, Inc.
40. Grameenphone Ltd. Bangladesh. http://grameenphone.com, accessed on 15.12.2012
41. Kritikal Solutions Ltd., India. http://www.kritikalsolutions.com/products/traffic-analyzer.html, accessed on 15.12.2012
42. Yang Q. and Koutsopoulos, H. N., (1996). A microscopic traffic simulator for evaluation of dynamic traffic management systems, *Transportation Research C, 4(3)*,113-129
43. Ben-Akiva M., Koutsopoulos H. N., Toledo T., Yang Q., Choudhury C. F., Antoniou C., and Balakrishna R. (2010). Traffic simulation with MITSIMLab, in Fundamentals of Traffic Simulation, 1st ed., ser. International Series in Operations Research and Management Science, J. Barceló, Ed. Springer, 233-268.
44. Population and Housing Census: Preliminary Results (2011), Bangladesh Bureau of Statistics, Statistics Division, Ministry of Planning, Government of the People's Republic of Bangladesh
45. Box M. J. (1965), A new method of constrained optimization and a comparison with other methods, *Computer Journal, 8(1),*42-52.