



UNIVERSITY OF LEEDS

This is a repository copy of *Development of origin–destination matrices using mobile phone call data*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/89507/>

Version: Accepted Version

Article:

Iqbal, MS, Choudhury, CF orcid.org/0000-0002-8886-8976, Wang, P et al. (1 more author) (2014) Development of origin–destination matrices using mobile phone call data. *Transportation Research Part C: Emerging Technologies*, 40. pp. 63-74. ISSN 0968-090X

<https://doi.org/10.1016/j.trc.2014.01.002>

© 2014 Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Development of Origin-Destination Matrices Using Mobile Phone Call Data

Md. Shahadat Iqbal¹, Charisma F. Choudhury^{1,2,3}, Pu Wang^{2,4}, Marta C. Gonzalez²

¹Department of Civil Engineering, Bangladesh University of Engineering and Technology, Bangladesh

²Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, USA

³Institute for Transport Studies, University of Leeds, UK

⁴School of Traffic and Transportation Engineering, Central South University, Hunan, P.R. China

Abstract

In this research, we propose a methodology to develop OD matrices using mobile phone Call Detail Records (CDR) and limited traffic counts. CDR, which consist of time stamped tower locations with caller IDs, are analyzed first and trips occurring within certain time windows are used to generate tower-to-tower transient OD matrices for different time periods. These are then associated with corresponding nodes of the traffic network and converted to node-to-node transient OD matrices. The actual OD matrices are derived by scaling up these node-to-node transient OD matrices. An optimization based approach, in conjunction with a microscopic traffic simulation platform, is used to determine the scaling factors that result best matches with the observed traffic counts. The methodology is demonstrated using CDR from 2.87 million users of Dhaka, Bangladesh over a month and traffic counts from 13 key locations over 3 days of that month. 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 (e.g.4,5,6,7,8). A number of Bayesian methods (e.g.9,10,11), Generalized Least Squares approaches (e.g.12,13,14), Maximum Likelihood Approaches(e.g.15), and Correlation Methods (e.g.16,17,18) 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 (e.g.19). More recent approaches for OD estimation include automated registration plate scanners (e.g.20) and mobile traffic sensors such as portable GPS devices (e.g.21,22,23). 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 (e.g.24). 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. 25,26,27,28), mobility pattern extraction (e.g. 29,30,31,32,33,34,35, 35), route choice modeling (e.g. 37,38), traffic model calibration (e.g. 39), traffic flow estimation (e.g.40,41) 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. (e.g.42) 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 Caceres et al. (e.g.43) 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 to 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. 44,45,29). Mellegård et al. (e.g.44) have developed an algorithm to assign mobile phone towers extracted from CDR to traffic nodes and Calabrese et al. (e.g.45) 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

1
2
3
4 have not been explored in detail. Wang et al. (e.g.29) have used an analytical model to scale up
5 the ODs derived from CDR by using the population, mode choice probabilities and vehicle
6 occupancy and usage ratios and have validated it using probe vehicle data. The methodology
7 however relies heavily on availability of traffic and demographic data in high spatial resolution
8 which may not be always available, particularly in developing countries.
9

10
11
12 In this research, we propose a methodology to develop OD matrices using mobile phone CDR
13 and limited traffic counts. CDR from 2.87 million users from Dhaka, Bangladesh over a month
14 are used to generate the OD patterns on different time periods and traffic counts from 13 key
15 locations of the city over a limited time are used to scale it up to derive the actual ODs using the
16 microscopic traffic simulator MITSIMLab. The methodology is particularly useful in situations
17 when there is limited availability of detailed travel survey and high resolution traffic data. The
18 ODs are validated by comparing the simulated and observed traffic counts of a different location
19 (which have not been used for calibration).
20
21
22

23
24 The rest of the paper is organized as follows. First we describe the data followed by the
25 methodology used for development of the OD matrix. The estimation and validation results are
26 presented next. We conclude with the summary of findings and directions for future research.
27
28

29 **2. Data**

30 2.1 Study Area

31
32
33 The central part of the Dhaka city has been selected as the study area and the major roads in the
34 network has been coded. This consists of 67 nodes and 215 links covering an area of about
35 300km² with a population of about 10.7million (e.g.46). The average trip production rate is 2.74
36 per person per day with significant portions of walking (19.8%) and non-motorized transport
37 trips (38.3%) (e.g.46).The traffic is subjected to severe congestion in most parts of the day, the
38 average speed being only 17km/hr¹.
39
40
41

42
43 The mobile phone penetration rate is approximated to be more than 90% in Dhaka (66.36%
44 being the national average) and Grameenphone Ltd. has the highest market share with 42.7m
45 mobile phone subscribers nationwide (e.g.47).
46
47

48 2.2 CDR Data

49
50 The CDR data, collected from Grameenphone Ltd, consists of calls from 6.9 million users (which
51 are more than 65% of the population of the study area) over a month. This comprises of 971.33
52 million anonymized call records in total made in between June 19, 2012 and July 18, 2012. The
53 majority of the users (63%) have made 100 calls or less over the month. The frequencies of users
54 making certain number of calls over the month and on a randomly selected day (15th July, 2012)
55
56
57
58
59

60 ¹Excluding the non-motorized vehicles which are restricted from entering the major roads
61
62
63
64
65

1
2
3
4 are presented in Figure 1. It may be noted that no demographic data related to the phone users
5 are available.
6

7 8 2.3 Traffic Count Data 9

10 Video data, collected from 13 key locations of Dhaka city network over 3 days (12th, 15th and
11 17th July 2012) have been used in this study to extract the traffic counts². The locations (shown
12 in Figure 2) have been selected such that they cover the major roads (links) of Dhaka city with
13 flows from major generators and governed by the availability of foot over bridges for mounting
14 video cameras. Care has been taken to avoid roads that have high percentages of non-motorized
15 transport and where lane-discipline is not strictly followed since simulation errors are likely to be
16 higher in these situations due to increased complexity of acceleration and lane-changing behavior
17 of drivers. The data has been collected for 8 hrs (8.00 am to 12.00 noon and 3.00 pm to 7.00pm)
18 and analyzed using the software TRAZER (e.g.48) to generate classified vehicle counts. Due to
19 inclement weather and poor visibility some portion of the data is non-usable though. Moreover,
20 TRAZER (which is the only commercial software that can deal with mixed traffic streams with
21 ‘weak’ lane discipline) has high misspecification rates in presence of high congestion levels and
22 in those cases, manual counting has been performed instead.
23
24
25
26
27
28
29

30 3. Methodology 31

32 Each entry in the CDR contains unique caller id (anonymized), the date and time of the call, call
33 duration and latitude and longitude of the Base Transceiver Station (BTS). A snapshot of the data
34 is presented in Figure 1. As seen in the figure, if a person traverses within the city boundary and
35 uses his/her phone from different locations that is captured in the CDR. CDR can thus provide an
36 abstraction of his/her physical displacements over time (Figure 3).
37
38
39

40 However, in the CDR data, a user’s location information is lost when he/she does not use his/her
41 phone. As shown in Figure 4, according to the CDR, a user may be observed to move from zone
42 B to zone E, but his/her initial origin (O) and final destination (D) may actually be located in
43 zone A and zone F. In such cases, segments of the trip information are unobserved in the CDR.
44 However, the mobile phone call records enable us to capture the transient origins and
45 destinations which may have the true origin and/or the destination missing for a portion of the
46 person’s itinerary, but still retain a large portion of the actual ODs. Thus, we use the concept of
47 transient origin destination (t-OD) matrix (as used by Wang et al. (e.g. 29), which uses the
48 mobile phone data to efficiently and economically capture the pattern of travel demand.
49
50
51
52

53 The second source of data used in this research is classified traffic counts extracted from video
54 recordings collected from 13 key locations of Dhaka. These counts represent the ground truth
55 but are more expensive to collect and limited in extent (only 3 days in this case). This limited
56
57
58
59

60 ²There are no loop detectors or any other automatic traffic counters in Dhaka
61
62
63
64
65

1
2
3
4 point source data therefore cannot be used as a stand-alone source to reliably capture the OD
5 pattern.
6

7
8 In this research, we therefore plan to combine the two data sources. The OD pattern is generated
9 using the CDR data and scaled up to match the traffic counts. The scaling factors are determined
10 using a microscopic traffic simulator platform MITSIMLab (e.g.49) using an optimization based
11 approach which aims to minimize the differences between observed and simulated traffic counts
12 at the points where the traffic counts are available.
13
14

15
16 The methodology is summarized in Figure 5 and described in the subsequent sections.
17

18 19 3.1 Generation of tower-to-tower transient OD matrix 20

21 The time-stamped BTS tower locations of each user are first extracted from the mobile phone
22 CDR data and used for generating tower-to-tower transient OD matrix. The CDR however
23 contains sparse and irregular records (e.g.33), in which user displacements (consecutive
24 non-identical locations) are often observed with long travel intervals i.e. the first location may be
25 observed at 8:56 and next location may be observed at 18:03 with no information about
26 intermediate locations (if any) or the time when the trip in between these two locations have been
27 made.
28
29
30

31
32 Another limitation of the CDR data is often there are changes in tower in the data in spite of no
33 actual displacement. This is because the operator often balances call traffic among adjacent
34 towers by allocating a new call (or shifting an ongoing call) to the tower that is handling lower
35 call volumes at that moment. In the CDR, such switches in towers are confounded with the actual
36 physical displacements. To reduce the number of false displacements and better identify timing
37 and origin-destinations of specific trips, we therefore extract displacements that have occurred
38 within a specific time window. A lower bound in the time window (10 minutes) is imposed to
39 reduce the number of false displacements without affecting the number of physical
40 displacements occurring within short intervals. An upper bound in the time window (1 hr) is
41 imposed to ensure that meaningful numbers of trips are retained. Therefore, a person trip is
42 recorded if in the CDR, subsequent entries of the same user indicate a displacement (change in
43 tower) with a time difference of more than 10 minutes but less than 1 hour.
44
45
46
47
48

49 Further, both call volumes (from CDR data) and traffic volumes (from traffic counts) had
50 significant variations throughout the day. Based on correlation analysis of total mobile call
51 volumes and total traffic counts (Figure 6), four time periods (7:00-9:00, 9:00-12:00,
52 15:00-17:00 and 17:00-19:00), have been chosen for analysis.
53
54

55 56 3.2 Conversion of tower-to-tower t-OD to node-to-node t-OD 57

58 For application of the t-ODs in traffic analyses, the origin and destination towers need to be
59 associated with corresponding nodes of the traffic network. The typical tower coverage area can
60
61
62
63
64
65

1
2
3
4 be represented as a combination of three hyperbolas (Figure 7), the size varying depending on
5 tower height, terrain, locations of adjacent towers and number of users active in the proximity
6 (which can vary dynamically).
7
8
9

10 The population density in the chosen study area is very high (more than 8111 inhabitants/sq. km
11 (e.g.53) and the tower locations are very close to each other (1 km on average). Because of the
12 high user density, it can be assumed that the area between two towers is equally split among the
13 two towers (Figure 8) that is, each tower t has a coverage area (A_t) approximately defined by a
14 circle of radius $0.5l$, where l is the tower-to-tower distance.
15
16
17
18

19 If a unique traffic node i overlaps with A_t , the calls handled by t are associated with node i (as in
20 the case of Tower 1 in Figure 6). However, if A_t has two (or more) candidate nodes for
21 association, then the candidate nodes are ranked based on the proportion of A_t feeding to each
22 node. That is, the node serving greatest portion of A_t is ranked 1, the node serving second highest
23 portion of A_t is ranked 2, etc. For example, in Figure 6, network connectivity (feeder roads) and
24 topography (presence of a canal with no crossing facility in the vicinity) denote that Node 1 and
25 Node 2 are candidate nodes for association with Tower 2. As the major portion of A_t is connected
26 to Node 2 and the remaining portion is connected to Node 1, they are ranked 1 and 2 respectively
27 for Tower 2. The data format after this step is presented in Figure 7b. As seen in the figure, this
28 typically consists of call records associated with unique nodes and in some cases, a few calls
29 associated with multiple candidate nodes. The calls are then sorted and ranked based on the
30 frequency of the unique nodes used by each user (based on analysis of his/her call locations over
31 the month). The frequency of occurrence of the candidate nodes are compared and used as the
32 basis of replacement. For example, frequency analysis of User “AAH03JA” indicates a higher
33 frequency of Node 1. Therefore, in cases where there are ambiguities between Nodes 2 and 1,
34 Node 1 is used (for this particular user). The tower-to-node allocations are thus user-specific
35 and uses information derived from each user’s overall travel pattern.
36
37
38
39
40
41
42

43 The same process is used for all users and node-to-node t -OD matrices for each time period of
44 each day are derived.
45
46

47 3.3 Finding the scaling factor and determining the actual OD matrix

48 As discussed, the node-to-node t -OD matrix (t - OD_{ij}) provides the trip patterns for developing the
49 actual OD matrix (OD_{ij}). However, in order to determine the actual OD matrix, the t -OD needs to
50 be scaled to match the real traffic flows. A scaling factor β_{ij} is used in this regard:
51
52

$$53 \quad OD_{ij} = \sum_{ij} (t-OD_{ij}) * \beta_{ij}$$

54
55
56 It may be noted that β_{ij} takes into account the market penetration rates (i.e. not every user has a
57 mobile phone or uses the specific service provider), the mobile phone non-usage issue (i.e.
58 mobile phone calls are not made from every location traversed by the user), the vehicle usage
59
60
61
62
63
64
65

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 (e.g.49) 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 (51,52) and freeway bias (see 50 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. It may be noted that MITSIMLab was calibrated prior to the OD estimation and the desired speed, acceleration and lane-changing model constants have been updated using detailed video trajectories (e.g.55,56,57).

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:

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

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

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

$t-OD_{ij,t}$ = 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:

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

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

Where,

$t-OD_{ij,t}^m$ = 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 is presented in Table 1.

Analyses of the node-to-node transientflows 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:

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

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

Where,

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

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

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

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 (e.g.54), the following values of scaling factors have been derived.

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.

1
2
3
4 The graphical representation of the t-ODs and actual ODs across the network for one of the time
5 periods and the variations for an example node are presented in Figures 10 and 11 respectively.
6
7

8 **5. Validation**

9

10
11 In addition to the aggregate data used for calibration, traffic counts are collected from four
12 additional locations on a different day (not used for calibration). For validation purposes, the
13 scaled up ODs have been applied to simulate the traffic between 9:00-12:00 in MITSIMLab and
14 the simulated traffic counts are compared against the observed counts from these locations. In
15 order to quantify the prediction error, Root Mean Square Error and Root Mean Square Percent
16 Errors have been calculated and are found to be 335.09 and 13.59% respectively.
17
18
19

20 **6. Conclusion**

21

22
23 The main outcome of this research is the methodology for development of the OD matrix using
24 mobile phone CDR and limited traffic count data. The strengths of both data sources are utilized
25 in this approach: the trip patterns are extracted from mobile phones and the ground truth traffic
26 scenario is derived from the counts. The methodology is demonstrated using data collected from
27 Dhaka.
28
29

30
31 There are several limitations of the current research though. Firstly, in this research a simplified
32 objective function with grouped scaling factors has been used. This overlooks the heterogeneity
33 in call rates from different locations (e.g., more calls may be generated to and from railway
34 stations compared to and from offices with land telephone lines, etc.). A more detailed
35 classification of scaling factor can be used to overcome this bias and may yield better results.
36 Moreover, in this particular context, detailed network data and extensive calibration data were
37 not available which limited the number of traffic nodes used in the study. The transferability of
38 the driving behavioral models in MITSIMLab have also not been tested in detail and only the
39 key behavioral model constants have been updated to better match the traffic patterns in Dhaka
40 (e.g.55,56). These factors may have increased the simulation errors and affected the validation
41 results. However, initial validation results indicate promising success in real life application by
42 transport planners and managers. It may be noted that though MITSIMLab has been used in this
43 study to determine the scaling factors, the developed ODs are simulator independent.
44
45
46
47
48

49
50 Since CDR is already recorded by mobile phone companies for billing purposes, the approach is
51 more economic than the traditional approaches which rely on expensive household surveys
52 and/or extensive traffic counts. It is also convenient for periodic update of the OD matrix and
53 extendable for dynamic OD estimation. This method is particularly effective for generating
54 complex OD matrix where land use pattern is heterogeneous and asymmetry in travelling pattern
55 prevails throughout the day but there is a limitation of traditional data sources.
56
57
58
59
60
61
62
63
64
65

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, the National Natural Science Foundation of China and the New England UTC.

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. Chungcheng Lu, Xuesong Zhou, Kuilin Zhang.Dynamic origin–destination demand flow estimation under congested traffic conditions.*Transportation Research Part C: Emerging Technologies*, Volume 34, 16-37, September 2013.
7. Leonardo Caggiani, Michele Ottomanelli, Domenico Sassanelli.A fixed point approach to origin–destination matrices estimation using uncertain data and fuzzy programming on congested networks.*Transportation Research Part C: Emerging Technologies*, Volume 28, 130-141, March 2013.
8. Nie, Y., H. M. Zhang and W. W. Recker (2005) Inferring origin-destination trip matrices with adecoupled GLS path flow estimator. *Transportation Research Part B*, 39, pp. 497 - 518.
9. Maher, M. (1983). Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research Part B*, 20(6), 435–447.
10. 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.
11. Li, B. (2005). Bayesian inference for origin–destination matrices of transport networks using the EM algorithm. *Technometrics* 47 (4), 399–408.
12. 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.
13. Bell, M. (1991). The estimation of origin–destination matrices by constrained generalized least squares. *Transportation Research Part B*, 25 (1), 13–22.
14. Toledo T. and Kolehkina T., Estimation of dynamic origin-destination matrices using linear assignment matrix approximations, *IEEE Transactions on ITS* 14(2), pp. 618-626, 2013.
15. Spiess, H. (1987). A maximum likelihood model for estimating origin-destination matrices, *Transportation Research Part B*,21(5), 395-412.

16. Vardi, Y. (1996). Network tomography: estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, 91, 365–377.
17. Hazelton, M.L. (2000). Estimation of Origin–Destination matrices from link flows on uncongested networks. *Transportation Research Part B*, 34(7), 549–566.
18. Hazelton, M.L. (2003). Some comments on origin–destination matrix estimation. *Transportation Research Part A*, 37(10), 811–822.
19. Hazelton, M.L., 2001b. Inference for origin–destination matrices: estimation, reconstruction and prediction. *Transportation Research Part B*, 35(7), 667–676.
20. 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.
21. 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.
22. 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.
23. 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.
24. <http://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2013.pdf> [accessed 20July, 2013]
25. 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.
26. 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.
27. 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.
28. YasuoAsakura, EijiHato. Tracking survey for individual travel behavior using mobile communication instruments. *Transportation Research Part C*. Volume 12, Issues 3–4, 273-291, June–August 2004.
29. Wang, P., Hunter, T., Bayen, A. M., Schechtner, K., and González, M. C. (2012). Understanding Road Usage Patterns in Urban Areas. *Scientific reports*, 2.
30. González, M. C., Hidalgo, C. A., and Barabási, A. L.(2008).Understanding individual human mobility patterns, *Nature*, 453, 779–782.
31. Song, C, Koren, T, Wang, P, and Barabási, A. L. (2010). Modelling the scaling properties of human mobility, *Nature Physics*, 6, 818–823.
32. 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.
33. 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.

- 1
 - 2
 - 3
 - 4
 - 5
 - 6
 - 7
 - 8
 - 9
 - 10
 - 11
 - 12
 - 13
 - 14
 - 15
 - 16
 - 17
 - 18
 - 19
 - 20
 - 21
 - 22
 - 23
 - 24
 - 25
 - 26
 - 27
 - 28
 - 29
 - 30
 - 31
 - 32
 - 33
 - 34
 - 35
 - 36
 - 37
 - 38
 - 39
 - 40
 - 41
 - 42
 - 43
 - 44
 - 45
 - 46
 - 47
 - 48
 - 49
 - 50
 - 51
 - 52
 - 53
 - 54
 - 55
 - 56
 - 57
 - 58
 - 59
 - 60
 - 61
 - 62
 - 63
 - 64
 - 65
34. 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.
35. Asgari, F., Gauthier, V., and Becker, M. (2013). A survey on Human Mobility and its applications. arXiv preprint arXiv:1307.0814.
36. Francesco Calabrese, MiDiao, Giusy Di Lorenzo, Joseph Ferreira Jr., Carlo Ratti. Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C. Volume 26*, 301-313, January 2013.
37. 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.
38. 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.
39. Bolla, R., Davoli, F., and Giordano, A. (2000). Estimating road traffic parameters from mobile communications. In *Proceedings 7th World Congress on ITS*, Turin, Italy.
40. 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.
41. Cheng P, Qiu Z, Ran B. Particle filter based traffic state estimation using cell phone network data. *Proceedings of the IEEE ITSC 2006*, 2006.
42. 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.
43. 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.
44. 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.
45. Calabrese F., Lorenzo G. D., Liu L. and Ratti C. (2011). Estimating Origin-Destination Flows using Mobile phone Location Data. *IEEE Pervasive Computing*, vol. 10(4), 2011, pp. 36-43.
46. 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.
47. Grameenphone Ltd. Bangladesh. <http://grameenphone.com>, accessed on 15.12.2012
48. Kritikal Solutions Ltd., India. <http://www.kritikalsolutions.com/products/traffic-analyzer.html>, accessed on 15.12.2012
49. 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
50. 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*, Volume 145 of *International Series in Operations Research and Management Science*, J. Barceló, Ed. Springer, 233-268.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

51. Hillel Bar-Gera. Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel. *Transportation Research Part C*. Volume 15, Issue 6, 380-391, December 2007.

52. Xianyuan Zhan, Samiul Hasan, Satish V. Ukkusuri, Camille Kamga. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C*. Volume 33, 37-49, August 2013.

53. Population and Housing Census: Preliminary Results (2011), Bangladesh Bureau of Statistics, Statistics Division, Ministry of Planning, Government of the People's Republic of Bangladesh.

54. Box M. J. (1965), A new method of constrained optimization and a comparison with other methods, *Computer Journal*, 8(1), 42-52.

55. Iqbal S. (2013), Development of Origin-Destination Trip Matrices Using Mobile Phone Call Data, MSc Thesis, Bangladesh University of Engineering and Technology.

56. Islam M. M. (2013), Acceleration Decision in Heterogeneous Traffic Stream, MSc Thesis, Bangladesh University of Engineering and Technology.

57. Siddique A. M. (2013) Lateral Movement Models for Heterogeneous Traffic Stream, MSc Thesis, Bangladesh University of Engineering and Technology.

Figure

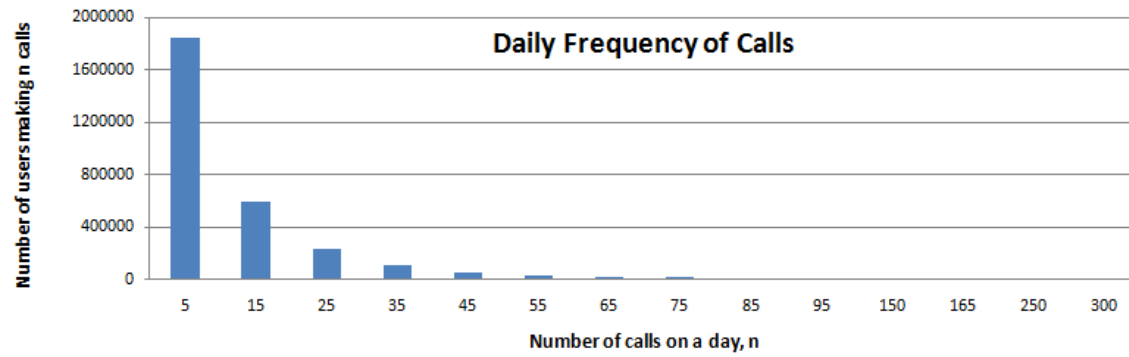
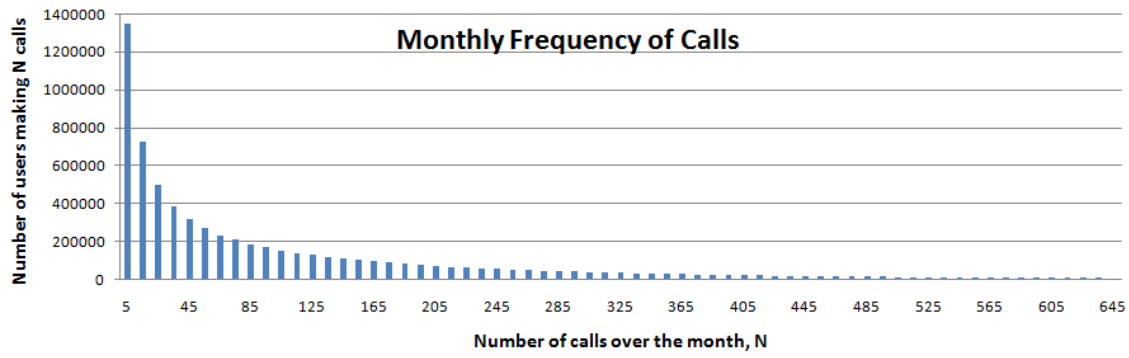


Figure 1: Frequency of calls per user

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

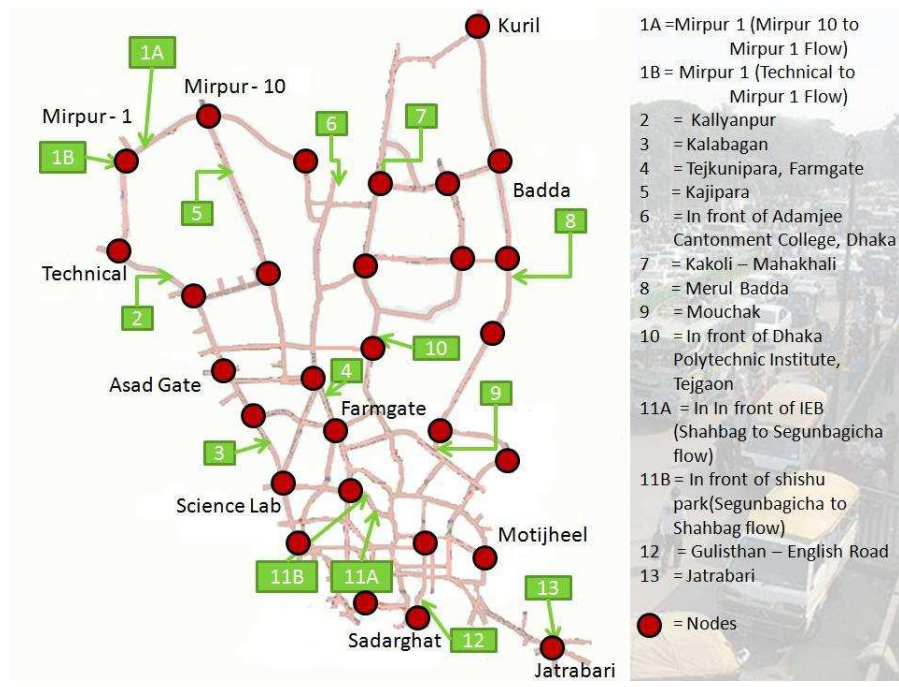


Figure 2: Locations of video data collection and position of OD generating nodes

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

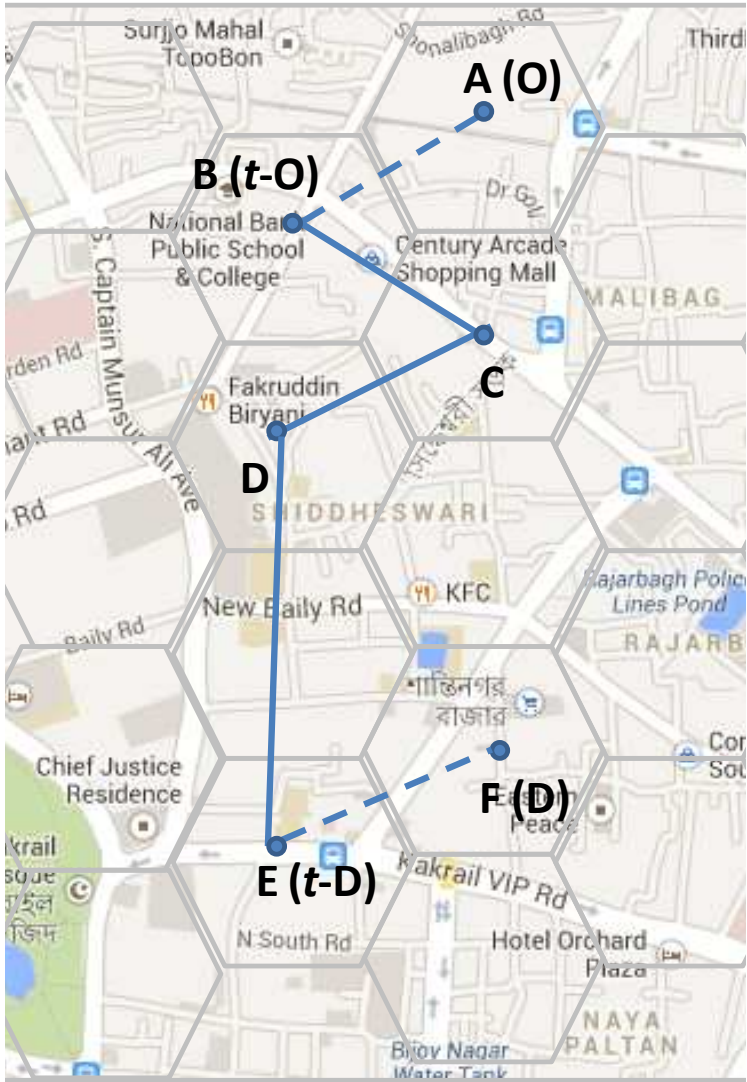


Figure 4: Actual vs. Transient OD

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

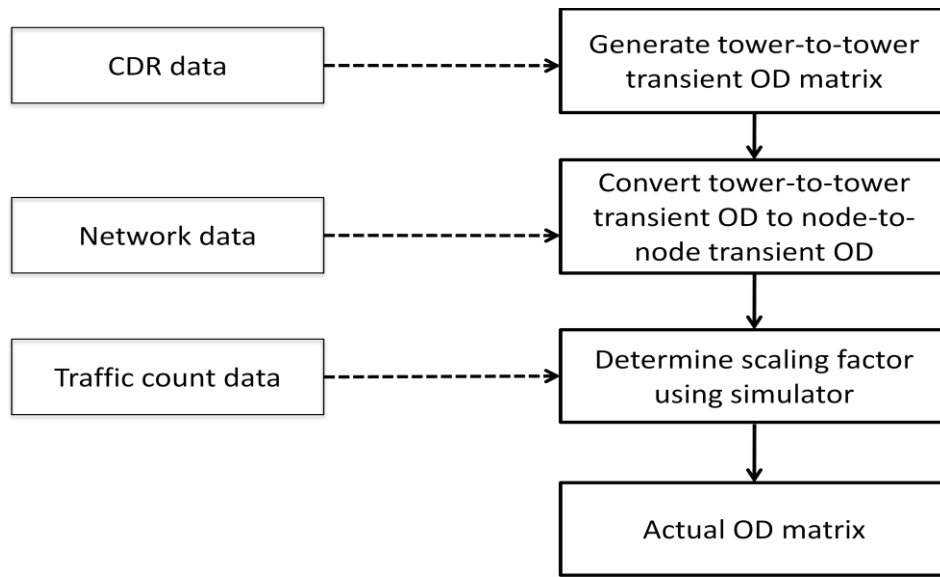


Figure 5: Framework for developing OD Matrix

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

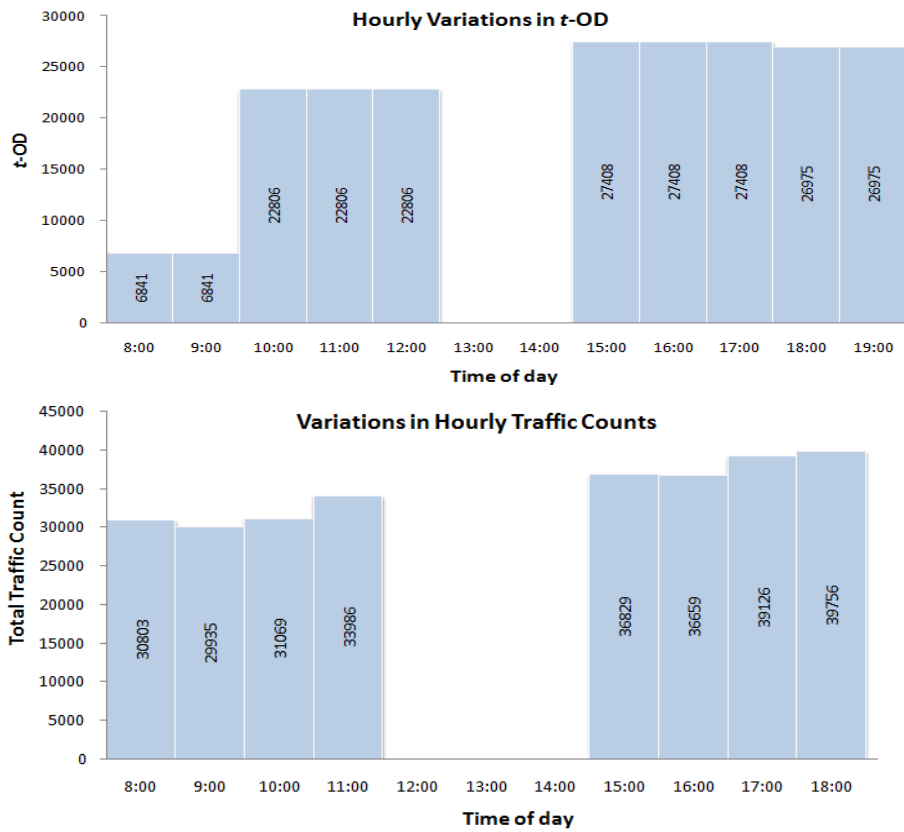


Figure 6: Hourly variations a. traffic count b. transient ODs from mobile call records

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

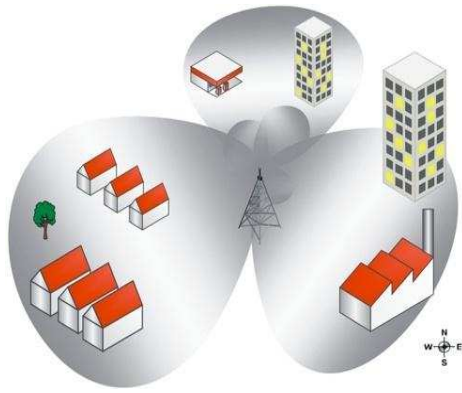
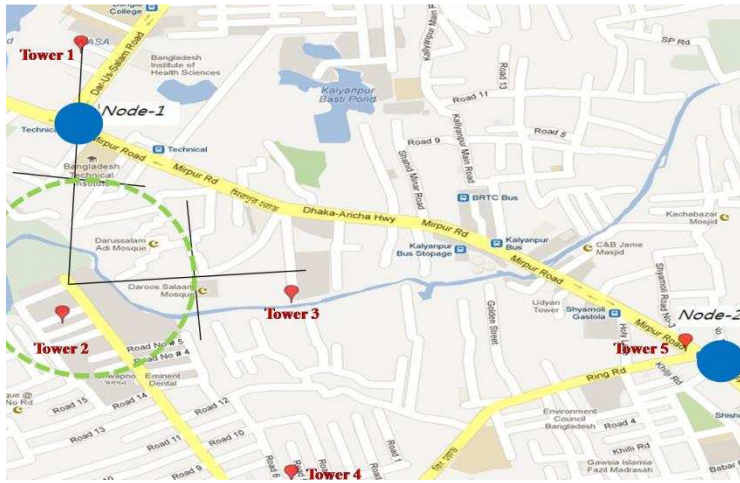


Figure 7: Typical coverage area of a tower (http://www.truteq.co.za/tips_gsm/)



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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

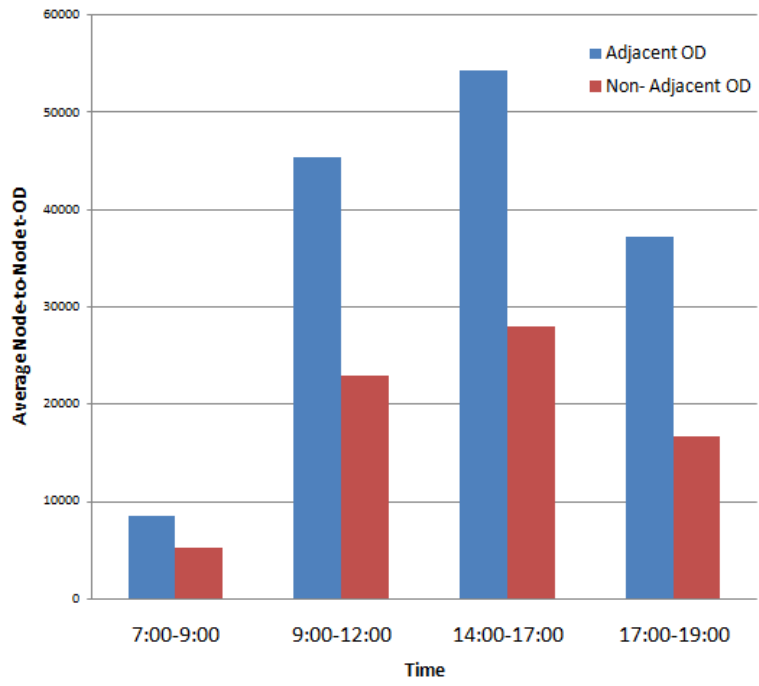
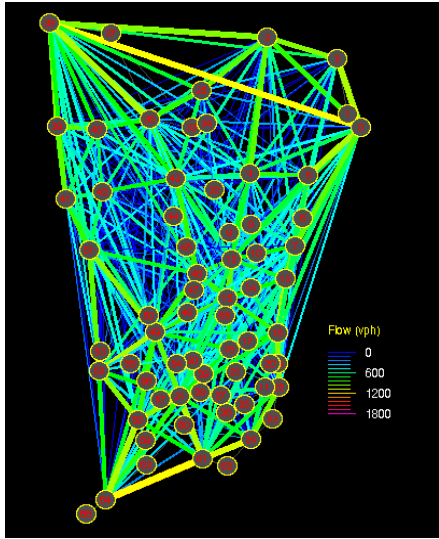
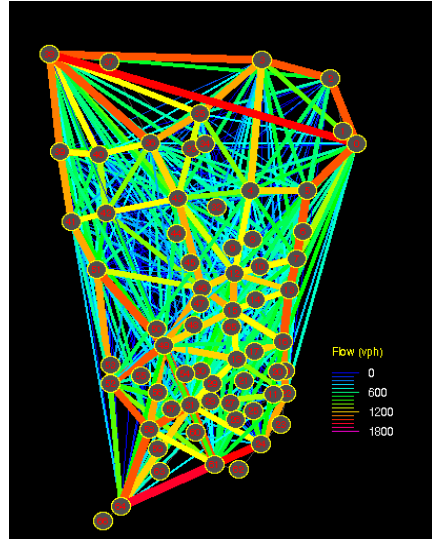


Figure 9: Comparison of t-ODs between adjacent and non-adjacent nodes

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65



a. t-OD



b. actual OD

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

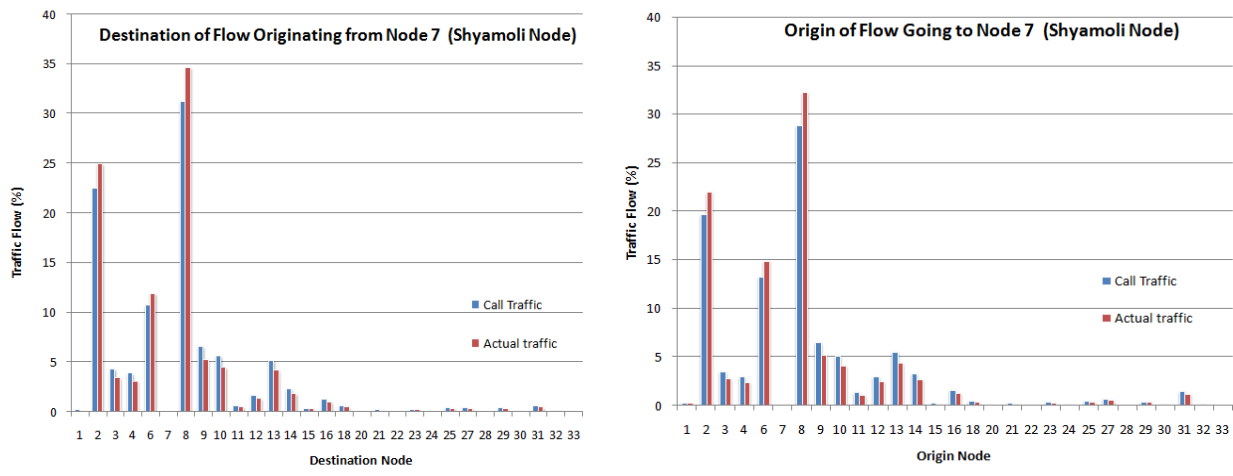


Figure 11: Example of Transient and Actual Traffic Flows To and From a Node (Shyamoli) between 7:00-9:00.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65**Table 1:** Node-to-node t-OD

Time Period	Time	t-OD	
		Total Over the Month (including weekends)	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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

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