

**UNDERSTANDING TRAVEL TO WORK ATTRIBUTES
USING MOBILE NETWORK BIG DATA
(STUDY AREA: WESTERN PROVINCE)**

N. K. Bhagya Jeewanthi
(168046N)

Thesis submitted in partial fulfillment of the requirements for the
degree of Master of Science

Department of Transport and Logistics Management

University of Moratuwa
Sri Lanka

February 2018

DECLARATION OF ORIGINALITY

I declare that this is my own work and this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Signature:

Date:

COPY RIGHT STATEMENT

I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date:

STATEMENT OF THE SUPERVISOR

The above candidate has carried out the research for the Degree of Master of Science under my supervision.

Name of the supervisor:

Signature of the supervisor: Date:

ABSTRACT

As people become more mobile, urban traffic patterns become more complex, creating a need for more continuous transportation planning processes. Currently, manual and online surveys are the primary source for such analysis. However, such data collection while being expensive, takes time and is often outdated by the time it is made available for analysis. Mobile Network Big Data (MNBD) which concerns large data sets has the potential to supplement such traditional data sampling programs. Call Detail Records (CDR) which is a subset of MNBD is readily available as most of the telecommunication service providers maintain such data. Thus, analyzing CDR leads to an efficient identification of human behavior and location.

This research uses the CDRs of nearly 10,000 mobile phone users in the Western Province (WP) of Sri Lanka for a period of three months for the analysis of their caller locations in order to determine their mobility patterns. In analyzing the CDRs, the frequency of making calls from a specific location is identified, classified them into potential home and non-home locations based on the regularity and time of day and week these calls were generated from each such location. Users are thereby categorized hierarchical levels based on the regularity of presence within the study area, identification of province of residence and by typical employment categories across the sample of users.

An estimate of home-based work trips made within the Western Province was identified using the CDR and validated by comparing with the origin-destination matrix for work trips calculated from an extensive survey of 35,000 households under the CoMTrans Study (JICA, 2014) obtaining a fit of 76%. The successful validation and the identification of sources of errors in CDR data provides direction for further research in using CDRs for travel estimation and the identification of the appropriate comprehensive data mining techniques.

Key words – Mobile Network Big Data, Call Detail Record, Home-based Work trips

ACKNOWLEDGEMENT

Prima Facie, I would like to offer my heartfelt gratitude to my research supervisor, Senior Prof. Amal S. Kumarage. I have been fortunate to have a supervisor like him who guided me from start to the end; in the meantime, followed up my progress continuously and guided me with a bunch of advices in each step of the research. I am very grateful for your encouragement and continuous attention throughout the research.

Special thanks should go to the post-graduate research coordinator Dr. Varuna Adhikariwattage and former coordinator Dr. Indika Sigera, for conducting research progress reviews and the support given through the process.

I would like to express my profound gratitude to Dr. T. Sivakumar, former Head of the Department of Transport and Logistics Management of University of Moratuwa and the current Head of the department. Mr. S. N. Bentotage for the opportunity created.

A very special gratitude goes to Sriganesh Lokanathan (Team leader, LirneAsia), Chairman of the review panel for lending me the data. I also thank Isuru Jayasooriya, Danaja Maldeniya (Researchers - LirneAsia) including Sriganesh Lokanathan who immensely contributed in the stages of primary data processing and analyzing. Specially thank you all for keeping trust on me, in providing a data and the guidance given from start to the end. I gratefully acknowledge the funding received towards my M.Sc. from the University of Moratuwa by the SRC grant (SRC/ST/2016/14) for “Mobility studies from mobile big data for Colombo”.

I am sincerely thankful to all the staff members and the non-academic staff of the Department of Transport and Logistics Management. I am also thankful to my family members, colleagues, senior colleagues, and friends who supported me in many ways throughout the research. Encouragements, advices, and opinions of yours are really appreciated.

TABLE OF CONTENTS

DECLARATION OF ORIGINALITY	ii
COPY RIGHT STATEMENT	iii
STATEMENT OF THE SUPERVISOR.....	iv
ABSTRACT	v
ACKNOWLEDGEMENT.....	vi
LIST OF FIGURES	x
LIST OF TABLES	xi
LIST OF ABBREVIATIONS	xii
CHAPTER 1 - INTRODUCTION.....	1
1.1 Background	1
1.2 Introduction to basic transport concepts.....	2
1.2.1 Identifying trips.....	2
1.2.2 Transportation forecasting	4
1.3 Introduction to Mobile Network Big Data	7
1.4 Purpose of the research.....	8
1.5 Research gap.....	9
1.6 Research Objectives	10
1.7 Data sources	12
1.8 Scope of the research.....	12
1.9 Overview of the thesis	13
CHAPTER 2 - LITERATURE REVIEW.....	15
2.1 Mobile positioning data.....	15
2.2 Use of Call detail records in transportation.....	16
2.3 Use of CDR in Sri Lankan context.....	20
2.4 Limitations of using CDR in Travel predictions	21
2.5 Techniques used in minimizing the limitations of CDR	23
2.6 Research gap identification	24

CHAPTER 3 – INTRODUCTION TO HVS DATA USED FOR RESULT	
VALIDATION.....	26
3.1 Introduction to Household Visit Survey.....	26
3.2 Methodology of the household visit survey	26
3.3 Trip purpose composition.....	27
3.3.1 Distribution of HBW trips.....	30
3.3.2 Hourly fluctuations of home-based work trips	31
3.4 Discussion	32
CHAPTER 4 – ANALYSIS OF CALL DETAIL RECORDS	33
4.1 Introduction	33
4.2 Methodology	34
4.2.1 Data description	34
4.2.2 Methodological framework.....	35
4.2.3 Value creation with Call Detail Records.....	37
4.3 A - User extraction based on caller activity threshold.	38
4.4 B - Identification of user Presence	39
4.4.1 Presence location identification	39
4.4.2 Time parameter	43
4.4.3 Categorization based on the presence in the Western Province.....	44
4.5 C - Identifying work and home locations	44
4.5.1 Aggregating presence of users	46
4.5.2 Minimizing the errors of CDR	48
4.5.3 Overview of the sample	49
4.5.4 Framework for identifying work – home distribution.....	50
4.5.5 User categorization based on the residential area.	52
4.5.6 User Model implementation based on the presence pattern of the day.	52
4.6 Summary of user categorization.....	56
CHAPTER 5 – VALIDATION OF WORK-HOME DISTRIBUTION.....	58
5.1 Introduction	58
5.2 Result validation.....	58
5.3 Statistical analysis	58

5.3.1	Validation of workplace distribution	59
5.3.2	Validation of home distribution	60
5.4	Conclusion.....	62
CHAPTER 6 –SUMMARY & CONCLUSION		64
6.1	Research Summary	64
6.1.1	Methodological framework of using CDR.....	64
6.1.2	Application of CDR based findings	65
6.1.3	Validation with HVS data	66
6.2	Limitations of the study.....	66
6.3	Future Research	67
REFERENCES		69
APPENDIX A		73
APPENDIX B		74
APPENDIX C		75
APPENDIX D		77
APPENDIX E		80
APPENDIX F.....		81
APPENDIX G		82

LIST OF FIGURES

Figure 1.1: Concept of a trip	3
Figure 1.2: Home based work trips	4
Figure 1.3: Four step model	4
Figure 1.4: Characteristics of residential and non - residential zones.....	5
Figure 1.5: Trip Estimation with CDR.....	10
Figure 3.1: Composition of trips by trip purpose	28
Figure 3.2: Hourly distribution of home based work trips.....	31
Figure 4.1: Methodological framework of the analysis.....	35
Figure 4.2: Initial user identification based on the nature of the records.....	39
Figure 4.3: Defining a presence	39
Figure 4.4: Example of user location identification_Gampaha DSD	40
Figure 4.5: Identifying administrative areas from tower locations	41
Figure 4.6: Cell towers within the Gampaha DSD	42
Figure 4.7: Initial user identification based on the nature of the records.....	43
Figure 4.8: Aggregated caller activity pattern of users across the sample of 5758 ...	45
Figure 4.9: Illustration of the core working/ home stay period	46
Figure 4.10: Illustration of the spatial distribution of a single user	47
Figure 4.11: Overview of the sample	50
Figure 4.12: Identifying work - home distribution.....	50
Figure 4.13: Summary_user categorization	57

LIST OF TABLES

Table 1.1: Types of trip movements.....	6
Table 1.2: Research direction.....	10
Table 1.3: Identified trip types	13
Table 2.1: Methodologies of using CDR in transport aspects	17
Table 2.2: Trip identification in stay based approach	21
Table 3.1: Estimation of the target sample size	27
Table 3.2: Description of the purposes	28
Table 3.3: Occupation types and the corresponding percentages	29
Table 3.4: Distribution of home based work trips.....	31
Table 4.1: Sample of data records- single user data after pre - processing.....	37
Table 4.2: User behavior illustration – absence of records in the 3rd month	38
Table 4.3: Cell coverage areas	42
Table 4.4: Example_ cumulating the total no.of days within the Western Province .	43
Table 4.5: Division of time categories	46
Table 4.6: Presence of a single user	47
Table 4.7: Basic user categorization according to user appearance.....	49
Table 4.8: Home work distribution _mobile data	51
Table 4.9: Home work ditribution_HVS data	53
Table 4.10: User model description	543
Table 4.11: Description of time categories	544
Table 4.12: Demonstration of the models	514
Table 5.1: Time window selection in validation of home locations	511
Table 5.2: Statistical resluts of two time categories.....	61

LIST OF ABBREVIATIONS

MNBD	Mobile Network Big Data
CDR	Call Detail Records
VLR	Visitor Location Register
CA	Caller Activities
HVS	Household Visit Survey
WP	Western Province
DSD	Divisional Secretariat Divisions
GND	Grama Niladhari Divisions
BTS	Base Transceiver Station
O- D MATRIX	Origin – Destination Matrix