ISSN: 2513-2520

R4TLI Conference Proceedings 2017
Paper Reference: *R4TL (17) – A33*

# Understanding Travel Behaviour from Call Detail Records

**N K Bhagya**
University of Moratuwa, Sri Lanka

**Amal S Kumarage**
University of Moratuwa, Sri Lanka

**Sriganesh Lokanathan**
LIRNEasia, Sri Lanka

## 1. Introduction

Understanding human mobility is essential in many fields including transportation planning. Currently, manually carried out surveys are the primary source for such data and analysis. Such data collection, while being expensive, takes time and is often outdated by the time it is available for analysis. Mobile Network Big Data (MNBD) has the possibility of supplementing traditional data sampling programs, resulting not only in cost and time savings, but with a striking growth in the amount of information available for analysis.

Mobile Network Big Data (MNBD) concerns large-volume, complex, growing data sets derived from the way people use communication devices. Compared to other network-related data like GPS, Call Detail Records (CDRs) are the largest subset of MNBD and are easily available since most telecommunication service providers maintain such data for billing purposes (Manoranjan Dash, 2015). Whenever a cellular transaction is made, a CDR which consists of time-stamped tower locations with caller IDs are generated (Md.Shahadat Iqbal, 2014). CDR describes the mobile usage pattern of a particular user of which the present focus is to extract information on the mobility of the user. Thus, analysing CDR will lead to understanding human mobility.

Moreover, CDRs have been used in various analyses which address different transport concepts. These include O-D trip estimation with different techniques (M. K. D. T. Maldeniya, 2016), identifying significant locations visited by users, and deriving home and work points through more advanced models developed for the purpose. However, in Sri Lanka there is less research on understanding travel behaviour. A firm theoretical account will enable the effective use of MNBD for transport demand forecasting.

## 2. Methodology

This study focuses on understanding the behavioural patterns of mobile users within the study area i.e., the Western Province of Sri Lanka. This analysis begins with the study of the appearance of different users and the consistency of such appearance within the Western Province. This is followed by the study of the regularity of user appearance within particular zones (also called stays). The study of the regularity of such stays over time is used to identify potential stays such as home and non-home locations.

The data set used for analysis consisted of a sample of randomly selected CDR of 1,000 IDs from across Sri Lanka, as obtained by one mobile operator over a period of three months. The analysis excluded callers who did not demonstrate regular mobile usage patterns while the remaining CDRs were processed to determine user appearances in a meaningful manner. The locations or stays of each user were identified by considering the appearance threshold within particular time categories. The connectivity between these locations were analysed further to derive significant travel patterns.

The output of mobility analysis based on MNBD can be aligned with the different stages of existing transport demand estimation techniques such as trip generations, attractions, trips between origin zones and destination zones, the identification of trip purposes and even route identification for regular travel.

### 2.1. Identifying Stays in Western Province

The CDR of the 1000 IDs were analysed by four time-categories, i.e.; weekday mornings, weekday evenings, weekend mornings and weekend evenings. The Provinces each user was found to have stayed in during the considered period were identified. Of them, users who had at least a single stay within the Western Province were extracted. Based on that 49% of users were qualified for further analysis. This can also be extrapolated to be understood that 49% of mobile callers would visit the Western Province at least once during a three months' period. The same analysis also helps us understand the frequency of visits by these callers.

### 2.2. Identifying Stay Locations in the Western Province

The records with any stay in the Western Province were then identified by the clustered location or zone defined by the tower locations within which such calls were identified. The study of the time patterns of these stays were also used to identify the probable activity of the caller within the zone such as home and non-home locations. For example, a caller who is regularly found during weekday mornings in a particular zone with regular stays in another zone during nights and on weekends will lead to the identification of a home zone and a non-home zone, while the trip is likely to be a home-based work or school trip.

## 2.3. Identifying Home and Non-Home locations

Home and other non-home locations of users were decided based on the total number of appearances of each user in each zone. Due to differences in mobile usage, it is difficult to define the exact appearance threshold to confirm a location as a home or a non-home location at this preliminary stage. Therefore, the study assumed that a caller who appeared to be found in a zone more than 75% of the time was considered to be a regular visitor to that specific zone within the considered time period. The most frequent zones during evenings were considered as the home location while regularly-visited zones during mornings were considered as non-home locations.

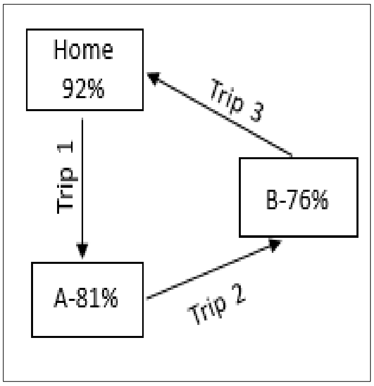


Figure 1 shows a typical caller stay within different zones in the Western Province, which identifies a specific location as home since 92% of the stays were at night or in the weekend, while locations A and B appear to be in the vicinity of the caller's home but were non-home locations since they were found to have been visited during 81% and 76% of the time respectively during weekday mornings.

**Figure 1: Example of movement between home and non-home locations**

However, many callers did not demonstrate clear home or non-home locations. These could be those who do not have regular trips away from homes, or those who have both their homes and offices or schools located in the same zone. Also itinerant persons such as salesmen, transport workers and tradesmen would also not demonstrate regular trip patterns.

## 2.4. Identifying Travel Purposes

Considering the regularity and the connectivity of regular stays, three user categories were identified: low mobility users, commuters and expatriates. When the most frequently used zone within the mornings and the evenings were the same, such callers were considered as low mobility users. Commuters are those whose frequent stay in the mornings is different to that of the evenings. These are regular commuters. Those who were found to frequent three top locations between weekday mornings, weekday evenings and weekends were identified as expatriates who stayed the week in temporary locations and regularly went to their permanent residences over weekends. The analysis was able to further identify those who were expats in the Western Province and those who were from the Western province. Results from the above caller types are shown in Table 1.

**Table 1: Summary of Caller Types (by percentage of users)**

<table>
<tr><th></th><th>Low Mobility Users</th><th>Commuters</th><th colspan="2">Expatriates</th></tr>
<tr><td>From 490 IDs</td><td>63%</td><td>20%</td><td colspan="2">17%</td></tr>
<tr><td rowspan="2">Primary stay locations within WP</td><td rowspan="2">52%</td><td rowspan="2">58%</td><td>Weekdays in WP</td><td>Weekends in WP</td></tr>
<tr><td>40%</td><td>60%</td></tr>
</table>

### 2.5. Validation with Household Visit Survey data

The research concludes its findings by validating the CDR results with Household Visit Survey data in the Western Province collected from extensive surveys conducted by the CoMTrans Study (JICA, 2014). Table 2 shows one such validation used to compare the percentage of those with clearly distinguished work and home locations within the Western Province, which amounted to 26% of the 223 residences, with the finding of the HVS survey that, of the total population of 5,821,710 in the Western Province, 27.9% or 1,624,632 people worked within the same province.

**Table 2: Validation with HVS data**

| Criteria | HVS data | CDR data - (sample of 1000 Caller IDS) |
|---|---|---|
| % of users with work and home location within Western Province | = (1624632/5821710) * 100%<br>**= 27.9%** | = (58/223) * 100%<br>**= 26%** |

## 3. Conclusion

The main outcome of this research is the identification of a methodology to understand and validate travel behavioural patterns within the Western Province using mobile phone CDR. The development of the methodology is demonstrated using 1,000 randomly selected IDs.

There are several limitations of the current research, primarily load-balancing caused by mobile devices oscillating between alternative towers indicating physical movement of a user even when the device and the user remain still. This results in the recording of false movements. Also, users who work on time shifts may be misclassified. A user who habitually makes several calls in route between stays may also have a large number of stays in zones which are actually transit zones. The pursuit of further analyses using larger CDR sample sizes will lead to better understanding of travel behaviour.

## 4. References

[1] Manoranjan Dash, “CDR-To-MoVis: Developing A Mobility Visualization System from CDR”, IEEE,2015.

[2] MD. Shahadat Iqbal,“Development of origin-destination matrices using mobile phone call data”,ScienceDirect, 2014.

[3] M. K. D. T. Maldeniya,” An Assessment of Mobile Network Big Databased Insights for Transport Planning in Sri Lanka. R4TLI Conference Proceedings”, 2016.

[4] LIRNEasia - a policy and regulation think-tank, CoMTRans project-Oriental Consultants-2014

***Keywords:*** *mobile network big data, call detail record, load balancing*