Identifying Call Detail Records (CDRs) Affected by Load Sharing

Yashothara Shanmugarasa, Viren Dias, Lasantha Fernando
LIRNE*asia*, 12 Balcombe Place, Colombo 08, Sri Lanka
{yashothara, viren, lasantha}@lirneasia.net

I. Introduction

Pseudonymized CDR datasets are increasingly used as a ubiquitous data source to extract insights into human mobility, transportation, social networks and disaster response behaviors. The utility of these datasets is impaired by the load sharing effect - during peak network traffic, some calls are offloaded to a base transceiver station (BTS) alternative to the closest. Mobile network operators can triangulate locations more accurately using other data sources. However, in the absence of such datasets, the potential implications of load sharing on research must be addressed cautiously.

Existing literature briefly explores this issue; Isaacman et al. [1] handled load sharing by spatially clustering BTSs for each subscriber. In this paper we discuss several criteria that can be used to identify records affected by load sharing. We apply these criteria to an existing CDR dataset to show that load sharing is a prevalent issue, especially in urban areas.

II. Methodology

We used pseudonymized CDR data spanning a single month for approximately 2 million subscribers who had more than 50% of their activity in the Western Province of Sri Lanka to carry out this research. We selected this province due to its widespread urbanization.

We clustered BTSs by drawing a grid of 1 km^2 cells across the country and the position of the grid was determined by minimizing the cumulative distance between the centre of the cell and the BTSs contained within it. The centres of these cells were designated as the new locations for the BTSs contained within.

A trip constituted of consecutive records of a subscriber recorded by different BTSs. We explored different criteria to narrow down possible records affected by load sharing:

- A mean velocity exceeding 110 km/h.
- Between neighboring BTSs.
- A distance less than 5km, 10km, 15km, 20km, 25km or 25km and above.

Given the velocity criterion, we conducted analyses based on all possible combinations of the remaining two criteria to identify records potentially affected by load sharing.

III. Results

Fig. 1, derived from Table I, suggests that beyond a distance of 15 km the increase in distinct subscribers affected by load sharing is minimal. This indicates that the load sharing effect is only significant within this radius. The marginal increments beyond this distance can be attributed to the presence of cloned devices or infrequent cases of load sharing.

TABLE I
RESULTS OF APPLYING DIFFERENT CRITERIA TO THE DATASET

Distance	Only neighbor cells	Pct. of distinct subscribers	Pct. of candidate records
5	False	48.81	0.72
10	False	51.85	0.89
15	False	52.32	0.92
20	False	52.42	0.93
25	False	52.46	0.94
>25	False	52.56	1.00
5	True	43.39	0.53
10	True	45.26	0.60
15	True	45.39	0.61
20	True	45.40	0.61
25	True	45.40	0.61
>25	True	45.41	0.61

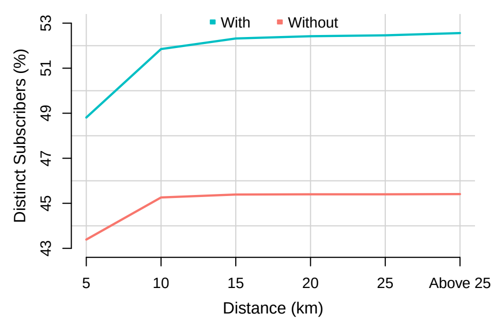

Fig. 1. Distinct subscribers affected by load sharing against distance

IV. Conclusion

Our method successfully captures a subset of records affected by load sharing. However, some of these records are the result of device cloning. The neighbor BTSs and distance criteria serve to isolate records affected by load sharing from those affected by device cloning, but not wholly. Further experimentation is required to identify a larger portion of records affected by load sharing and to segregate them from those affected by device cloning.

References

[1] Sibren Isaacman et al. "Identifying important places in people's lives from cellular network data". In: *International Conference on Pervasive Computing*. Springer. 2011, pp. 133–151.