FLOWMINDER.ORG

The impact of apparent ‘teleports’ on the estimation of mobility from Call Detail Records (CDRs)

Robert Eyre (robert.eyre@flowminder.org), Galina Veres, Véronique Lefebvre
Flowminder Foundation

Population mobility and call routing. The assumption underpinning the extraction of mobility information from Call Detail Records (CDRs) or other mobile operator data is that, if a subscriber is recorded from cells that are in different locations, then the subscriber has moved. This assumption is not always true however, as a subscriber’s consecutive calls can be routed by distant cells even if the subscriber has not moved. Call rerouting (or rerouting of SMS and data sessions) therefore creates spurious movements if a change in subscriber recorded location is interpreted as movement, which can make it appear as if subscribers travel further, and more often than they actually do.
Spurious movements can be detected when a change in recorded location occurs too quickly to be a plausible movement (e.g. 2 calls made with a 5 sec interval and routed by cells distant by 20km, which would be a speed of 14,400 km/h), which we call ‘teleports’. However, rerouting can also create spurious movements that are plausible (e.g. 2 calls made with a 5 hour interval and routed by cells distant by 20km) and not distinguishable from actual movements. Therefore we can only measure the tip of the iceberg (‘teleports’) of the larger ‘rerouting-without-movement’ problem.

Measuring and weighting ‘teleports’. To investigate the rerouting-without-movement problem we start by measuring its visible part (the teleports) and experiment with reducing their numbers through 1) spatial aggregation of recorded locations, and 2) weighing down the number of connections (recorded changes in subscriber location) observed for pairs of locations where teleports are common.
We used 7 months of CDRs from a random sample of 40,000 subscribers of Digicel Haiti and counted the number of subscribers’ consecutive connections for each pair of cell towers, and for each pair of level 3 administrative units (admin3). We then classified as ‘teleports’ all connections made at a speed exceeding 120km/h and computed the proportion of teleports over all connections, for each pair of locations.
Figure 1 shows the ratio of teleports to all connections, between each pair of cell towers (left) and each pair of admin3 units (right). For some pairs of towers none of the observed connections are real movements (all connections are ‘teleports’) and many pairs have high ratios of ‘teleports’. In comparison, aggregating towers by admin3 (which we commonly do in our mobility indicators) leads to lower proportions of ‘teleports’ as rerouting is more likely to occur over shorter distances and so more within admin3 units than across units. Grouping of towers near to administrative boundaries and taking cell coverage and direction into account would further reduce teleports (and all spurious movements).
We also observe that teleports are more likely in some locations, such as densely covered areas and areas where cells have longer coverage (e.g. coastal areas, flat terrain). Connections classified as teleports can be excluded from counts or ‘1 - proportion of teleports’ can be used to weight down aggregated numbers of connections for each pair of locations. This would need to be re-computed over time or averaged for specific time periods, as ‘teleports’ are more likely during the day or during events (e.g. new year’s eve) when the network is busy. Weighing down the number of connections using the proportion of teleports would improve comparisons of mobility across different locations, and over time, e.g. avoiding that mobility estimates are dominated pairs of locations with large proportions of ‘teleports’. However, ‘teleports’ are only the visible spurious movements and the number and proportion of all spurious movements are not accounted for.

Identifying subscribers’ ‘meaningful locations’ to attenuate the rerouting-without-movement influence. In addition to adjusting aggregated statistics we can also attempt to correct individual CDR trajectories by clustering nearby cells that are frequently recorded for each subscriber, to obtain a set of ‘meaningful locations’ for each subscriber. The assumption is that most rerouting-without-movement would occur within the meaningful locations, and that a change in meaningful location has a higher chance of being a real movement. We used the Hartigan’s Leading algorithm[1] to explore this method on our sample. **Figure 2** shows the cell towers used by a synthetic subscriber (left), how these towers would be clustered if only taking their location into account (middle), and the clusters obtained for this particular synthetic subscriber, i.e. their ‘meaningful locations’ (right). The proportion of teleports between meaningful locations is smaller than the proportion of teleports between fixed cell clusters (based on cell locations only), however when aggregating clusters by admin3 units the proportion of teleports is the same whether clusters were derived from individual CDR traces or from cell locations only, so

[1] Hartigan, John A. 1975. Clustering Algorithms. Wiley.

the identification of individual meaningful locations would only be worth computing for problems requiring sub-admin3 resolution.

Population mobility and call routing during crises. When the network is busy, increased rerouting is necessary, and to more distant locations, to maximise the chance for all subscribers to access the network. Increased phone usage and rerouting-without-movement may occur in crises (e.g. following disasters or during conflicts), which could lead to overestimates in crisis-driven displacements and erroneous ranking of most affected locations. This is particularly concerning as CDR-derived estimates could mislead response and recovery efforts by focusing on the wrong locations. Methods to correct for spurious movements (e.g. weighing by teleport proportions and identification of meaningful locations) would need updating immediately at the onset of a crisis and regularly throughout, to cope with fast changes in both network activity and mobility, and keeping with constraints of limited compute power and the requirement to provide updates in a timely manner.

Immediately upcoming work will investigate the rerouting-without-movement problem on the Digicel Haiti subscriber base during and after the 2021 earthquake in Haiti, and other recent displacement crises, to compare with routine patterns and and further test the envisaged correcting methods. Methods we implement operationally also need to allow for frequent (ideally daily) production of mobility estimates on a limited infrastructure at mobile network operators, to provide critical information to guide humanitarian and developmental efforts in the low and middle-income countries we work in.

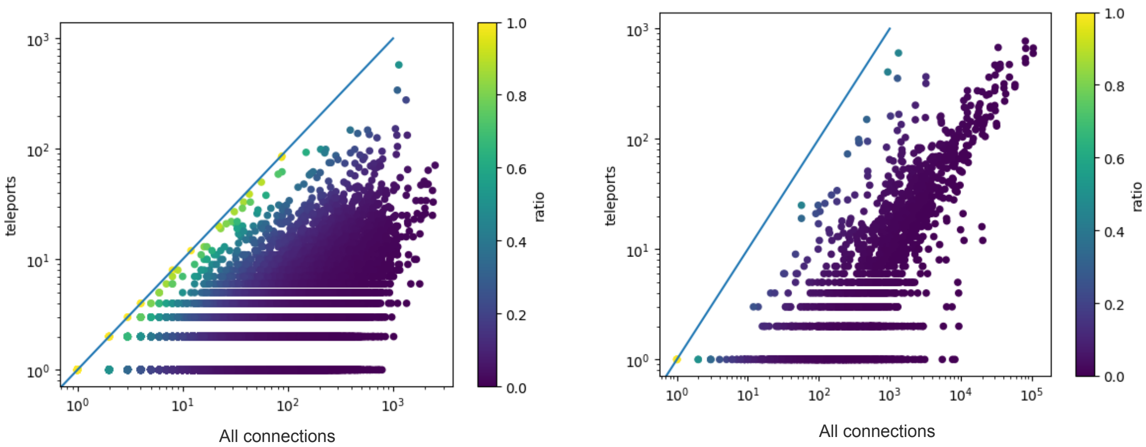

Figure 1: **Total teleports over the study period vs all connections for each pair of cell towers (left) and admin3 units (right)**. Large ratios indicate that the number of teleports is close to the number of connections (i.e. none or few of the recorded changes in location correspond to movements). The proportions of teleports between admin3 units are lower than between cell towers as rerouting is a local process.

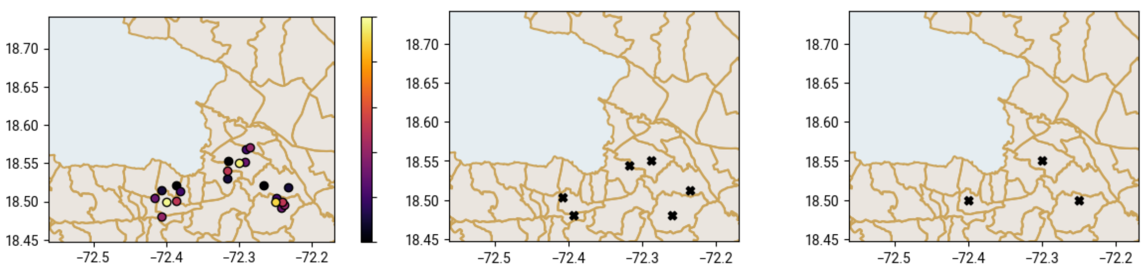

Figure 2: **Synthetic data: Identifying a subscriber's 'meaningful locations' to attenuate the rerouting influence. Left**: Synthetic example of the (randomly generated) cell towers serving a single (randomly generated) subscriber over a period of time, coloured by frequency of records. **Middle**: Example of clustering of cell towers based on their location - same clustering for all subscribers (crosses correspond to cluster centroids). With such clustering there is a risk to separate nearby locations that are commonly visited by a subscriber (or nearby cell towers that reroute their calls when they don't move), potentially resulting in an inflated number of trips between clusters. **Right**: Clustering of cell towers derived from the synthetic records of the (randomly generated) subscriber, identifying their 'meaningful locations' which aggregate both short movements and rerouting into a single spatial unit (which needs to be updated over time).