Inferring Urban Population Distribution from GSM Data: An Experimental Case Study

Hisham Raslan
Telecommunications Industry Consultant
Teradata Egypt
Giza, Egypt
hisham.raslan@teradata.com

Ahmed Elragal
Business Informatics & Operations Department
German University in Cairo (GUC)
New Cairo, Egypt
ahmed.elragal@guc.edu.eg

***Abstract*-A huge amount of location and tracking data is gathered by tracking and location technologies, such as Global Positioning system (GPS) and Global System for Mobile communication (GSM) devices leading to the collection of large spatiotemporal datasets and to the opportunity of discovering usable knowledge about movement behavior. Movement behavior can be extremely useful in many ways when applied, for example, in the domain of planning metropolitan areas, traffic management, mobile marketing, tourism, etc. Currently, population statistics are derived from census data and household surveys. However, they are difficult to manage and costly to implement. In this paper, we move towards this direction and propose a technique for inferring urban population distribution from GSM data. A case study is used from to build a prototype for testing and evaluating the proposed technique. Results showed that we were able to infer population density with high accuracy based on analyzing GSM data.**

Keywords-Population density; GSM-CDR data; experimental case study.

I. INTRODUCTION

Spatiotemporal patterns that show the cumulative behavior of a population movement can be very useful in understanding mobility-related phenomena. In fact, the discovery of the pattern of traffic flows among sequences of different places in a town can help decision makers to take well informed decisions in different areas, such as urban planning and traffic management. In many application domains, useful information can be extracted from moving object data if the meanings as well as the background information are considered. The knowledge of moving patterns between different places in the geographic space can help the user answer queries about moving objects or movement behavior.

The mainstream literature focuses on finding movement behavior using GPS datasets, however, GPS is rarely used in some developing countries, e.g., Egypt. On the other hand, almost overwhelming majority of the population is using mobile phones (in Egypt, >90M lines which exceed 100% penetration). For this reason, in this research we move towards the direction of using GSM data, as an alternative to the unavailable GPS data.

In this paper, we introduce a technique for identifying home and work locations based on GSM-CDR (Call Detail Records) data. CDR stores the details of the call (caller ID, called ID, time of the call, call duration, etc. Before applying the technique, POI (Points Of Interest) will be identified, and then, the identified POI will be mapped to CDR data, which would then be used in inferring the location. The research relies on experimental case study based on GSM data from a mobile operator to support answering the research question "*how to infer population density from GSM-CDR data*?"

The remaining of this document is organized as follows: Section II provides an overview of related work. Section III describes the proposed framework, conceptual model, and the algorithm used. In Section IV, the proposed framework is applied on an experimental case study. Section V discusses the results of the experimental case study. Section VI evaluates the work done and recommend future research.

II. RELATED WORK

The wide use of technologies, such as Global System for Mobile communication (GSM) networks and Global Positioning System (GPS) resulted in the availability of large amounts of spatiotemporal data. This section provides an overview of related work within the domain of spatiotemporal data analysis and mining research community including related work and research areas that directly address inferring population density using spatiotemporal data.

The identification of dense areas has received the attention. Nevertheless, the techniques used so far suffer from limitations including poor spatial resolution due to the use of grids. Vieira et al. [1] proposed the DAD-MST algorithm to identify dense areas from CDR that is able to process large scale datasets and respects the original tessellation of the space. The DAD-MST algorithm has been tested and validated using a real CDR dataset of almost 50 million entries for over 1 million unique users over a four-month period. The dense areas identified have been qualitatively validated using the subway system of the city under study. The dynamics of the dense areas identified revealed the use that the citizens make of their city, indicating differences between different hour ranges and weekdays and weekends.

One billion people now live in an urban slum, the vast majority of them in developing nations [2]. A slum can be defined as “a residential area which has developed without

 ISBN: 978-1-61208-383-4

legal claims to the land and/or permission from the concerned authorities to build; as a result of their illegal or semi-legal status, infrastructure and services are usually inadequate". Sociologists theorize that the majority of urban migration is filtered through slums and understanding the migration patterns is vital to understanding the growth of urban areas. Wesolowski and Eagle [2] used data generated from mobile phones to better understand one of the largest slums, Kibera located in Nairobi, Kenya; focusing on migration patterns out of Kibera, inferring places of work, and tribal affiliations. Kibera has one cell tower location inside the slum, identified with six unique cell tower IDs. In order to form a sample of Kibera's residents, a caller is classified as living in Kibera if they meet all of the following criteria: 1) Over fifty percent of their total calls between the hours of 6 PM and 8 AM have been made from one of Kibera's towers. 2) The total number of calls made in a month is between 3 and two standard deviations from the mean number of calls made by those living in Kibera. They used this identification of callers, looking at three key components of slum dynamics: migration trends, work trends, and tribal affiliations.

Understanding the causes and effects of internal migration is critical to the effective design and implementation of policies that promote human development. Typically government censuses and household surveys do not capture the patterns of temporary and circular migration that are prevalent in developing economies. Blumenstock [3] illustrated how new forms of information and communication technology can be used to better understand the behavior of individuals in developing countries and used mobile phones as a new source of data on internal migration. Using Rwanda as a case study, they developed and formalized the concept of inferred mobility, and computed this and other metrics on a large dataset containing the phone records of 1.5 million Rwandans over four years. The empirical results corroborate the findings of a recent government survey that notes relatively low levels of permanent migration in Rwanda. However, this analysis reveals more subtle patterns that were not detected in the government survey, namely, high levels of temporary and circular migration and significant heterogeneity in mobility within the Rwandan population. The following statistics are computed based on the mobile phone transaction history: Number of cell towers used: As a very crude proxy for the movement of the individual, it is the count of unique towers used by the individual during the specified interval of time. Maximum distance travelled: This is the maximum distance between the set of towers used by the individual over the interval under study. Radius of Gyration (ROG): measures how far an object travels from its center of gravity. In the case of humans, ROG roughly measures the typical range of a user in space.

People spend most of their time at a few key locations, such as home and work. Being able to identify how the movements of people cluster around these "important places" is crucial for a range of technology and policy decisions in areas such as telecommunications and transportation infrastructure deployment. Isaacman et al. [4] proposed a new technique based on clustering and regression for analyzing anonymized cellular network data to identify generally important locations, and to discern semantically meaningful locations such as home and work. The algorithm used for identifying important places has two stages. In the first stage, it spatially clusters the cell towers that appear in a user's trace. In the second stage, it identifies which of the clusters are important using a model derived from a logistic regression of volunteers' CDR's. To select Home or Work, the relevant algorithm (i.e., either the Home or Work algorithm) calculates a score for each important cluster using coefficients obtained from a logistic regression. The algorithm then assigns the cluster with the highest score to be Home or Work.

Understanding the spatiotemporal distribution of people within a city is crucial to many planning applications. Obtaining data to create required knowledge, currently involves costly survey methods [5]. In their research, Toole et al. examined the potential of using GSM-CDR data to measure spatiotemporal changes in population. In the process, they identified the relationship between land use and dynamic population over the course of a typical week. A machine learning classification algorithm was used to identify clusters of locations with similar zoned uses and mobile phone activity patterns. It was shown that the mobile phone data is capable of delivering useful information on actual land use that supplements zoning regulations.

In their attempt to perform a comparative analysis of the behavioral dynamics of rural and urban societies, Eagle et al. [6] estimated the location based on cellular towers. They used four years of mobile phone data from all 1.4M subscribers within a small country. They did not have access to phone numbers, but rather unique IDs that provide no personally identifiable information. In addition to the standard information within CDR, which includes voice and text-messages.

To understand dynamics of human mobility in support of urban planning and transportation management, Phithakkitnukoon et al. [7] developed an activity-aware map that contains most probable activity associated with a specific area in the map based on POIs information. With activity-aware map, they were able to extract individual daily activity patterns from analyzing a large mobile phone data of nearly one million records.

Palmer et al. [8] conducted a pilot study - the Human Mobility Project (HMP) – to explore the use of mobile phones in demographic research and to test a technique: a dynamic, location-based survey. The pilot study uses Global Positioning System (GPS) and cellular tower data from mobile phones to determine subjects' residence locations, observe how they respond to questions at a fixed time and place and also to examine where they spend time when not at home, their trajectories, and how they respond to questions at a variety of different times and places. They were able to recruit 270 volunteers in 13 countries, to share GPS and cellular tower information on their trajectories and respond to dynamic, location-based surveys using an open-source Android application.

 ISBN: 978-1-61208-383-4

The methods currently available for monitoring traffic tend to require the installation of ancillary devices along roadways (loop detectors, cameras and so on), which, along with the costs of installation and maintenance, render alternative techniques more attractive [9]. Their research project has demonstrated a methodology to infer traffic data, such as journey Origin Destination matrices or traffic counts at given points in the road network using a technique that involves analysis of anonymous phone location data in a mobile phone network. The estimation of traffic data has been carried out using non-real data generated by a simulator of vehicular traffic and phones along a stretch of the road network. These simulated data comprise phone location data, which form the source for the development of this technique.

Smoreda et al. [10] reviewed several alternative methods of collecting data from mobile phones for human mobility analysis and described cellular phone network architecture and the location data it can provide. In conclusion, the authors proposed considering cellular network location data as a useful complementary source for human mobility research and provided case studies to illustrate the advantages and disadvantages of each method.

Liao et al. [11] introduced a hierarchical Markov model that can learn and infer a user's daily movements through an urban community. The model uses multiple levels of abstraction in order to bridge the gap between raw GPS sensor measurements and high level information such as a user's destination and mode of transportation. Locations, such as bus stops and parking lots, are learned from GPS data logs.

Krumm [12] assessed the privacy threats and countermeasures associated with location data. They examined location data gathered from volunteer subjects to quantify how well four different algorithms can identify the subjects' home locations and then their identities using freely available, programmable web search engine. Their procedure was able to identify a small fraction of the subjects and a larger fraction of their home addresses. Then, they applied three different obscuration countermeasures designed to foil the privacy attacks: spatial cloaking, noise, and rounding and showed how much obscuration is necessary to maintain the privacy of all the subjects

Advances in sensor networking and location tracking technology enable location-based applications but they also create significant privacy risks. Gruteser et al. [13] proposed a distributed anonymity algorithm that is applied in a sensor network, before service providers gain access to the data. The purpose of these mechanisms is to provide a high degree of privacy, save service users from dealing with service providers' privacy policies, and reduce the service providers' requirements for safeguarding private information.

III. PROPOSED TECHNIQUE

In our research, we consider population density is a result of residential and business facilities in areas (or POI) and in order to estimate the density we need first to identify the home and work locations for every commuter (user). The question is "how to identify home and work POIs for commuters from the CDRs?"

Every time a mobile subscriber makes or receives a call the network generates a CDR to store the details of the call (caller ID, called ID, time of the call, call duration, etc.). One of the CDR parameters is the cell id, which can be used to identify cell location. It is possible to develop an algorithm to find home and work locations for the commuters based on categorizing day of the week and hour of the day to home, work, travel, and other. Then we aggregate the number of CDRs created by the commuters in different locations at the defined categories; then the location where we find the user most at a specific category, home or work, is considered the commuter home or work POI. Following are the steps to identify commuters home and work POI:

1. Define "location type" as Home, Work, Travel, or Other;
2. "Location Type" value is based on time of the day and day of the week as suggested in Table I;
3. Aggregate the number of CDRs done by commuters for each Location Type per location (POI);
4. For home and work Location Types, the POI with the maximum number of CDRs for a specific location type is assigned as the commuter home or work POI respectively;
5. Knowing home and work POI, we can aggregate to the required level (district, area, city, and governorate).

TABLE I. DEFINITION OF LOCATION TYPE

Day of Week		Time Interval		Location Type	Time Interval Description
Start	End	Start	End		
1	7	0	6	Home	Home-overnight (Sun - Sat)
6	6	6	11	Home	Home-Friday morning
2	5	20	24	Other	Other-Afternoon Free time
6	6	11	24	Other	Other-Rest of Fri
7	7	0	24	Other	Other-all day, Sat
1	1	0	24	Other	Other-all day, Sun
2	5	6	8	Travel	Travel-Morning rush hour
2	5	16	20	Travel	Travel-Afternoon rush hour
2	5	8	16	Work	Work-regular working hours

In order to apply this technique, we need to identify the mobile operator cells covering the areas under investigation then map their locations to geographical areas (semantics); then, using the CDRs generated with the identified cells over a period of time we can identify operator subscriber's densities at these areas.

We propose a framework that aims to establish an enhanced approach towards building analysis engine that can infer population density from GSM data. The framework includes process, technologies, data, and decisions as the main aspects towards knowledge extraction.

The phases of the proposed framework are presented in Figure 1.

The process outlines the detailed tasks to take place at each phase and highlights the main layers which data passes through to be transformed into knowledge.

 ISBN: 978-1-61208-383-4

The proposed framework process is comprised of the following phases: storage, data preparation, data pre-processing, analytics, and interpretation. The framework takes all phases into consideration to provide a holistic view of the solution.

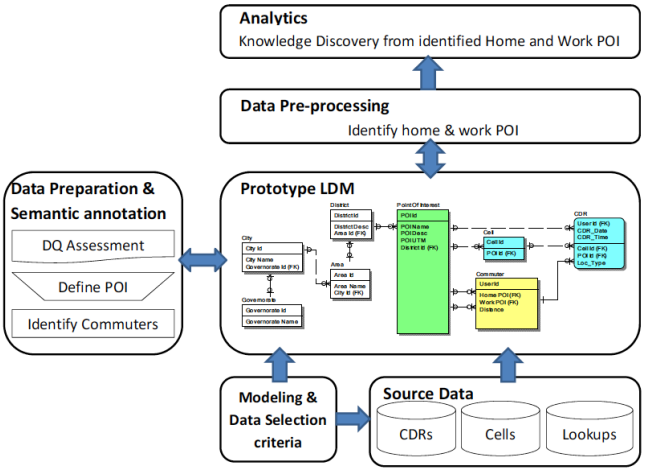

Figure 1. Our Framework.

While using GSM data to study traffic pattern, we will not be violating people's privacy, we will be using disguised data, which do not reveal real identify of people. Besides, unlike GPS data which identify exact location of a subscriber, a GSM cell covers a wide area that makes it impossible to identify the exact location of a subscriber by knowing the location of the cell used during the network activity. This limitation will not affect our research as we are interested in the collective movement behavior between areas.

The following are the tasks performed to reach the results:

1. Build Logical Data Model (LDM);
2. Data loading
   - Load CDRs and Cell information
   - Create and load lookups
3. Data preparation & Semantic annotation
   - Data Quality assessment
   - Define POI to add semantics to the data
   - Identify Commuters (users)
4. Infer Home and Work POI;
5. Perform Analytics.

IV. EXPERIMENTAL CASE STUDY

In this section, we apply the suggested framework on an experimental case study. Our case study is based on data collected, with permission, from a mobile operator in Egypt. The data represents 4 months of CDR's describing Greater Cairo.

A. Build the Logical Data Model

The foundation for the required analysis is the Logical Data Model (LDM). The LDM is the conceptual design that shows the business entities and relations between them. It is used to design the Physical Data Model (PDM), which will be created physically on the database management system (DBMS). The LDM is used to help the analytical users plan and develop queries and analytics. Figure 2 shows the LDM used for the case study.

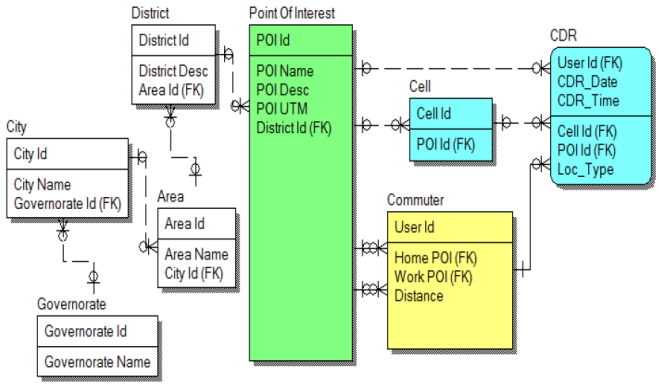

Figure 2. Prototype LDM.

The model is comprised of entities, attributes and relations; the model entities are colored to differentiate the type of information stored in each entity; blue entities are data loaded from the source (CDRs and Cells), white (uncolored) entities have semantics information (lookups: governorate, cities, areas, districts), green represents entities identified for analysis (POIs), and yellow entities contain information extracted from the data (Commuters). The LDM is used to create the PDM, which will be created physically on the selected DBMS.

B. Data Loading

Data loading includes two main tasks: (i) loading CDRs and cell information; and (ii) creating and loading lookups. Load CDRs and Cell information is accomplished by the following activities:

- CDR tables for voice and GPRS are merged into the table "CDR". The CDR table has the following structure: (User_Id, CDR_Date, CDR_Time, and Cell_Id)
  - Four Months of voice and data CDR's are merged into the table "CDR".
  - Total number of CDRs: 10,314,009,634 ≈ 1.5B
  - Greater Cairo cells loaded to "Cell" table in the cell id column; the POI column will be mapped later.
  - Number of cells: 14,175
- After loading the CDR table, cell ids are loaded to the "Cell" table in the cell id column; the POI column will be mapped later. POI's were manually identified using Google earth by locating all cells on the map then grouping the cells in a specific area to a POI as we will explain later. The cell table has the following structure: (Cell_Id, POI_Id).

The create and load lookups fulfill the requirements of the data loading phase by creating those lookup tables:

Governorate, City, Area, and District. They are then uploaded with the information required to add semantics to the analysis results.

C. Data preparation

The first step in the data preparation id the data quality assessment. Data quality is the suitability of data to meet business requirements. Because different organizations and applications have different uses and requirements for the data, data quality requirements will also differ. So data do not have to be perfect, but they need to meet business requirements. We will be assessing the data for the following:

- Consistency (Format and Content)
- Completeness
- Uniqueness
- Integrity

We use data profiling tool to perform the required data quality assessments. Table II and Table III show value analysis for the CDR and Cell tables respectively.

TABLE II. VALUE ANALYSIS FOR THE CDR TABLE

Column Name	Count	null	Unique	Zero	Positive	Negative
User_ID	10314009634	0	15424023	0	10314009634	0
CDR_Date	10314009634	0	122			
CDR_Time	10314009634	0	86400	138790	10313870844	0
Cell_ID	10314009634	0	15077	0	10314009634	0

TABLE III. VALUE ANALYSIS FOR THE CELL TABLE

Column Name	Count	null	Unique	Zero	Positive	Negative
Cell_ID	14175	0	14175	0	14175	0
Cell_Status	14175	151	4			
POI_Id	14175	0	522	0	14175	0

In our experimental dataset, CDR's with cell ids that do not exist in the cell table represent:

1. Number of CDRs with missing cells: 214026321 (2% of total CDRs)
2. Number of missing cells: 1127 (8% of total Cells)

The ratios of discovered data issues (2%) will not affect the accuracy of the analysis results.

Second step in the data preparation is to define POI's. This step is comprised of the following:

1. Group cell sites to define POI: POI's under investigation are defined by nearest cell sites. To locate cell sites, all cells were added to Google earth. To add the cells to the map we use the site coordinates. The coordinates are in Universal Transverse Mercator notation (UTM). Cell sites usually have more than one cell and they all have same UTM. The process of grouping cells to a centralized POI is a manual task. We tried to choose the POIs to be close to a real POI and in the same time to identify the urban areas. This process reduced the number of points on the map from around 14K to about 500 points, which is more manageable and easier to locate on the map. The identified POI and the POI demographics are then loaded in the POI table.
2. Update POI column in the cell and CDR tables: As described earlier, the grouping of the cells to POI is done manually and a mapping table was created to map cells to POIs. The mapping table is used to populate the POI column in the cell table and by joining the cells table and the CDR's table on the cell id column we can update POI column in the CDR table.

Third and last step here is to identify commuters. Commuters are the users of the mobile devices who generate voice and data CDR's. To maintain the privacy of the subscribers we only have anonymous identification number (user id) for the subscribers in the CDR table. We can use user id to select the CDRs made by each user (subscriber) and use it for the purpose of this research. To populate the commuter table we select the distinct users from the CDR table and load them in the commuter table. The number of commuters (users) identified is 13,887,256 which is a considerable number of commuters, relative to greater Cairo population, that can fairly represent greater Cairo population distribution.

D. Infer Home and Work POI

Here, the proposed technique described earlier is applied to infer home and work POI in the commuter table. After populating the table, exploration queries are executed to assess the technique results. Resulting indicated that 71% of commuters where identified with home and work POI's (approximately 9.8M); 7% with POI only; 18% with Work only; and 4% with neither. Additionally, and in same process, approximately 10.8M home POI's were identified; 12.3 work POI's. 6.7M commuters were identified with identical home and work. After populating the table, we executed exploration queries to assess the technique results; the resulting figures of the queries are shown in Table IV.

TABLE IV. DATA EXPLORATION RESULTS

Commuters with Home_POI and Work_POI identified	9,831,746	71%
Commuters with Home_POI only identified	950,244	7%
Commuters with Work_POI only identified	2,463,754	18%
Commuters with neither Home_POI nor Work_POI	641,512	5%
T O T A L	13,887,256	100%
Identified Home_POI	10,781,990	78%
Identified Work_POI	12,295,500	89%
Unidentified Home_POI	3,105,266	22%
Unidentified Work_POI	1,591,756	11%
Commuters with identical Home_POI and Work_POI	6,796,530	49%

E. Perform Analytics

By aggregating per the required dimension (POI, District, Area, City, or Governorate) on Home or Work POI, then sorting the results descending, we were able to get the answer for different business question e.g., top busy cities as well as top busy governorates. In Table V, we provide results for Governorates as a sample.

 ISBN: 978-1-61208-383-4

TABLE V. TOP BUSY GOVERNORATES

Home (Governorate)				Work (Governorate)			
Id	Governorate	Total	%	Id	Governorate	Total	%
1	Cairo	5,361,306	39%	1	Cairo	6,444,982	46%
2	Giza	3,103,579	22%	2	Giza	3,365,887	24%
3	Qalubia	2,317,105	17%	3	Qalubia	2,484,631	18%
	Unidentified	3,105,266	22%		Unidentified	1,591,756	11%
	TOTAL	13,887,256			TOTAL	13,887,256	

The numbers are inferred from four month of usage data from June till September 2012. The inferred numbers can be related to the actual numbers as follows:

- The inferred populations of the residence of Cairo, Giza and Qalubia governorates as shown in the table above are about 5M, 3M, and 2M respectively
- According to Egypt census bureau, the population of Cairo, Giza and Qalubia governorates are about 9M, 7M, and 5M respectively [14]
- The percentages for the inferred population relative to the announced are: 55%, 42%, and 40% and these are reasonable percentages for the following considerations:
    - The data used are coming from one operator out of three operators in Egypt;
    - The operator form which we have obtained the data, has been the first to start its operations in Cairo which enabled it to acquire the largest customer base among the three operators;
    - Mobile penetration in Egypt by Q3-2012 was about 113% [15].

As far as movement between cities is concerned, home_POI and work_POI were used to calculate the number of commuters moving between them and aggregating on the dimensions district, area, city, and governorate; we will be able to calculate the volume of traffic between each as follows:

- Select distinct home_POI, work_POI and Count of rows from commuter table to get the number of commuters moving daily from the two POIs
- Sort the results descending on the resulted count will show the two POIs with the highest volume of traffic
- Similarly, do the same to get governorates with the most traffic in and out.

A matrix that shows the movement between governorates is shown in Table VI.

TABLE VI. MOVEMENT BETWEEN GOVERNORATES

Home	WORK						
	Giza	%	Cairo	%	Qalubia	%	TOTAL Identified Home
Giza	2,447,841	87%	311,390	11%	43,562	2%	2,802,793
Cairo	214,168	4%	4,532,280	93%	146,419	3%	4,892,867
Qalubia	64,653	3%	261,946	12%	1,809,487	85%	2,136,086
TOTAL Identified Work	2,726,662		5,105,616		1,999,468		9,831,746

V. ANALYSIS OF RESULTS

In this paper, we presented a new technique for inferring population density from spatiotemporal data. The technique could be used in relation to various applications e.g., urban planning and traffic management. We have shown, conceptually and practically, how the technique was able to population distribution in greater Cairo and hence being able to help answer related business questions with regards to movement pattern between cities.

Results showed that it has been possible to infer the population density in areas, cities, and governorates with high accuracy. The analysis can also explain the regular movement volumes of commuters from home to work, which is very useful in identifying movement between areas, cities, or governorates and calculating the average distance of the commuters.

The proposed technique was able to detect movement patterns from high-granularity GSM data whereas mainstream literature focuses on finding these patterns from GPS fine-grained data.

VI. CONCLUSION AND FUTURE WORK

In this research, we proposed a new technique to infer population distribution and movement based on GSM-CDR dataset. While mainstream literature focus on finding this knowledge based on GPS data, we have been able to show how this could be achieved based on GSM data. In the case study, we used real GSM data that was provided by a mobile operator in Egypt. Our dataset included around 4 months of data for 13 million users performed over 10 billion calls and GPRS sessions. We performed different types of analyses covering population density and traffic between areas, cities, governorates. Urban planning and traffic management decision makers could use our technique to make related decision.

Our analysis confirms that long-term GSM activity data is well-suited typical population density analysis, especially when GPS data is not available.

Using real data in the case study was definitely for the benefit of the research. However, the data size was a major obstacle that caused the research to halt several times before Teradata granted the use of one of its servers to the research. This is very important to mention as GSM CDRs are always huge in volume and for deployment the required volume of data will be much more as data from all operators should be integrated to provide complete view for whole population.

Future work includes the deployment of our technique and in different business domains as well as conducting a comparative study between GSM versus GPS data.

REFERENCES

[1] M. R. Vieira, V. Frias-Martinez, N. Oliver and E. Frias-Martinez, "Characterizing dense urban areas from mobile phone-call data: Discovery and social dynamics," in IEEE Second International Conference, Minneapolis, Aug. 2010, pp. 241-248.

[2] A. P. Wesolowski and N. Eagle, "Parameterizing the Dynamics of Slums," in AAAI Spring Symposium: Artificial Intelligence for Development, Palo Alto, Mar. 2010, pp. 103-108.

 ISBN: 978-1-61208-383-4

[3] J. Blumenstock, "Inferring Patterns of Internal Migration from Mobile Phone Call Records:Evidence from Rwanda," Information Technology and Development, vol. 18 no. 2, 2012, pp. 107-125.

[4] S. Isaacman, R. Becker, R. Caceres and S. Kobourov, "Identifying Important Places in People's Lives from Cellular Network Data," in Proc. of 9th International Conference on Pervasive Computing (Pervasive), Berlin, 2011, pp. 133-151.

[5] J. L. Toole, M. Ulm, M. C. González and D. Bauer, "Inferring land use from mobile phone activity," in In Proceedings of the ACM SIGKDD International Workshop on Urban Computing, Beijing, Aug. 2012, pp. 1-8.

[6] N. Eagle, Y. de Montjoye and L. M. Bettencourt, "Community computing: Comparisons between rural and urban societies using mobile phone data," in CSE'09. International Conference, Vol. 4, Vancouver, Aug. 2009, pp. 144-150.

[7] S. Phithakkitnukoon, T. Horanont, G. Di Lorenzo, R. Shibasaki and C. Ratti, "Activity-aware map: Identifying human daily activity pattern using mobile phone data," in Human Behavior Understanding, Istanbul, 2010, pp. 14-25.

[8] J. R. Palmer, T. J. Espenshade, F. Bartumeus, C. Y. Chung, N. E. Ozgencil and K. Li, "New approaches to human mobility: Using mobile phones for demographic research," Demography, vol. 50, no. 3, 29 Nov. 2013, pp. 1105-1128.

[9] N. Caceres, J. P. Wideberg and F. G. Benitez, "Deriving origin–destination data from a mobile phone network," Intelligent Transport Systems, IET, 1(1), Mar. 2007, pp. 15-26.

[10] Z. Smoreda, A.-M. Olteanu-Raimond and T. Couronné, "Spatiotemporal data from mobile phones for personal mobility assessment," in Transport survey methods: best practice for decision making, London, 2013, pp. 1-20.

[11] L. Liao, D. J. Patterson, D. Fox and H. Kautz, "Learning and inferring transportation routines," Artificial Intelligence, vol. 171, no. 5–6, Apr. 2007, pp. 255-362.

[12] J. Krumm, "Inference attacks on location tracks," in In Pervasive Computing, Toronto, May 2007, pp. 127-143.

[13] M. Gruteser, G. Schelle, A. Jain, R. Han and D. Grunwald, "Privacy-Aware Location Sensor Networks," in HotOS, Vol. 3, Lihue, Hawaii, May 2003, pp. 163-168.

[14] Egypt census bureau, "Publication Name:Population," January 2013. [Online]. Available: http://www.capmas.gov.eg/pdf/EgyptInFigure /EgyptinFigures/Tables/English/pop/population/index.html. [Accessed Jan 2015].

[15] Egypt ICT Indicators, "September 2012 English.pdf," September 2012. [Online]. Available: http://www.egyptictindicators.gov.eg/en /Publications/PublicationsDoc/September%202012%20English.pdf. [Accessed Jan 2015].

 ISBN: 978-1-61208-383-4