Monitoring System of Urban Population Traffic based on Mobile Network Signaling

Jiangtao Luo, Zhongling Shu
School of Communication and Information Engineering
Chongqing University of Posts and Telecommunications
Chongqing 400065, China
Email: Luojt@cqupt.edu.cn; 13635432100@139.com

Yunfeng Zhou
Branch of Data Service
China Mobile Group Beijing Co., Ltd.
Beijing 100027, China
Email: zhouyunfeng@bj.chinamobile.com

***Abstract*—Online monitoring of population traffic within a certain region plays a significant role in modern city administration and public security, and it is also a difficult task to fulfil using traditional methods. In this paper, a system is reported, which has realized fast estimation of snapshot population and residence time distribution within a certain region, based on mobile network signaling. The system architecture is presented and some key implementation issues are addressed. As examples, results in two typical regions are demonstrated and thoroughly compared. One is a historical attraction, and the other is a traffic junction in the central business area. The system has currently come into operation in Beijing and begun serving several municipal departments including tourism, administration and public security.**

I. Introduction

In modern cities, it is very helpful to know the real-time population in a concerned region, the *regional snapshot population* (RSP). It will benefit many branches such as transportation system design, tourism prediction, administration management and even security guarding. In some cases, we need further to know how long people stay in a certain region and where they may come from. In order to evaluate the population in a region, we must be able to locate a person within a region, i.e. positioning. However, traditional positioning technologies cannot meet all requirements mentioned above. Therefore, we first reviewed related work on mobile user positioning technologies, and then introduced our system approach.

The main contribution of this paper was twofold. First, we presented a model to fast estimate the population and the residence time in a certain region, based on the discrete cell-based communication events. Second, we implemented the model in a monitoring system on a data warehouse, deployed it and made it a service for various industries and administration.

II. Related Work

There has been lots of work reported on locating the mobile phones with different precision and complexity. Choices are made depending on the application scenarios and trade-off between accuracy and costs. Generally, mobile positioning technologies can be divided into two kinds: GPS (global positioning system or its buddies) and non-GPS. Although most smart phones have been equipped with GPS, the majority of subscribers would not like to report their positions at all times. Therefore, passive monitoring technologies like non-GPS ones are more easier to utilize. Non-GPS positioning technologies can be further divided into mobile-based, network-based and hybrid, according to the place where the location is calculated. The simplest method of determining the mobile position is based on the Cell-ID, with the lowest precision with a range between 100 m and several kilo-meters. Some complex algorithms were proposed to improve the precise, like the AOA (Angle Of Arrival), TOA(Time Of Arrival) and TDOA (Time Difference Of Arrival) [1]. However, these techniques generally require software installation or modification on the mobile phone and additional location determination units and related software in the network side. Combining more than one method is another commonly used way to improve the precision [2]. The key issue is to measure the time or time difference from the mobile to multi-BS (base station). The measurements are implemented usually by analyzing the measurement reports at active mode [3]. Actually, the mobile location can be calculated by a third system. Related signaling messages are duplicated from the network and analyzed by a third application independent on the mobile and network. This is a completely passive solution, almost adding nothing to the phones and networks, and much easy to deploy. Such a system is appreciated by most network operators, because not all specifications in the latest standards have been implemented in all the network equipments and phones.

Based on mobile positioning, much research has been done on the road and urban traffic estimation [4] [5] [6], trajectory pattern mining [7] [8] and origination-destination (O-D) Analysis [9] [10], pushing and cultivating the ITS [11], LBS [12] and other unpredictable services and applications [13]. However, few work was done on estimating the regional snapshot population and average residence time within a

978-1-4799-4146-9/14/$31.00

closed area. In this paper, such a monitoring system will be reported and some typical results will be presented.

III. Principles

A. Problem Definition

The objective is to online estimate the population within a defined region and calculate the average time people stay in the region. Here, a **Region**, or a *Region of Concern* (RoC) is a geographical area defined by user, represented by a closed graph on the map.

A mobile phone is said to be in a region if an observed communication event reported it was served by the cell in the region, and no any event has reported that it has changed to another cell outside the region. The mobile communication events monitored include power on/off, MOC (Mobile Originated Call)/MTC (Mobile Terminated Call), mobile sent and received shot messages, location update (normal/periodical) and handover. The event making an outside subscriber become an inside one is called an Appear Event (AE). Accordingy, the event making an inside subscriber become an outsider is called a Leave Event (LE).

B. Algorithms

It is not self-evident to determine the current population within a region using the communication events since users are moving and the events are discrete. Before deciding whether a MS is in a RoC, we have to explore all the events related with it globally.

1) Calculation of Regional Snapshot Population hourly: According to the algorithm 1, the key is to determine the initial observed group, containing all people stayed in the region at the start of the period, and the gone group containing the population leaving the region and never returning during the period. All observed events formed a dataset of a sequence of time-spatial events of mobile users. The last event related with the MS is most significant, determining the final regional state of the MS. Therefore, the most time-consuming task is to search for the latest event.

2) Calculation of Residence Time: It is to obtain the distribution of people's residence time in a RoC. The residence time is classified into several stages in hours. For a certain statistical period (Δ), there are five scenarios of regional residence.

STAY-OUT case: i.e. the MS is observed to be outside the RoC after several inside events and never return again. The residence time in this period is the occurring time of last inside event minus the start point of the current period, which will be added to the old residence time of this MS, generating the final residence time.

IN-OUT case: i.e. the MS is observed to enter and then leave the region. The interval between the last and first inside event is regarded as the residence time. If there is only one inside event during the period, the first outside event is taken as the virtual last inside event.

IN-STAY case: i.e. the MS is observed to enter and never leave. The residence time is the time interval between the end point of the period and the occurring time of the first inside event. The end state of the MS is INSIDE, and its residence time will be accumulated in next period.

OUT case: i.e. the first event of an insider is an outside one. In this case, just change the regional state of the MS.

Duplicate OUT-IN case: i.e. the MS is observed entering and leaving this region for more than once time. It will be dealt with as several independent IN-OUT procedures.

Algorithm 1 Calculation of RSP

Require:
The concerned region, R_i;
The end hour, t_j;
Set of MSs in the region R_i at t_{j-1}, $\Phi_{R_i,t_{j-1}}$;
Set of events during the past hour, $E_{j-1,j}$.

Ensure:
RSP for the region of R_i at t_j, $RSP(R_i, t_j)$

```
Φ_r ← Φ_{R_i,t_{j-1}}, Φ_g ← ∅
  // Initialization: Φ_r contains the retained MSs on the hour; Φ_g contains gone MSs.
Select all MSs in E_{j-1,j} where region == R_i and event == APPEAR, forming Φ_{R_i,A};
  // Φ_{R_i,A} contains all MSs ever appeared in this region during the past hour.
Φ_o ← Φ_{R_i,t_{j-1}} ∪ Φ_{R_i,A}
  // Aggregate and derive the observed group.
for each s_k ∈ Φ_o do
  Search all the events of s_k in E_{j-1,j};
  E_{j-1,j,s_k} = {e_0, e_1, ...e_l}
    // e_l is the latest event.
  if e_l == LEAVE then
    Insert s_k in Φ_g;
  end if
end for
Φ_r ← Φ_o \ Φ_g
  // Erase gone MSs and then obtain the final retained population in Region R_i at t_j.
Count Φ_r, then get RSP(R_i, t_j).
```

IV. Monitoring Schema

A. System Overview

The system architecture is diagrammed in Fig. 1, which consists of four parts described as follows.

The signaling analysis (SA) part is to capture the raw data from the mobile network, decode the signaling messages and generate all kinds of *call detail record* (CDR), such as MOC, MTC, LU and etc.. Each CDR has several fields to describe its key attributes like time stamp, MS identity, LAC-CI and CDR type.

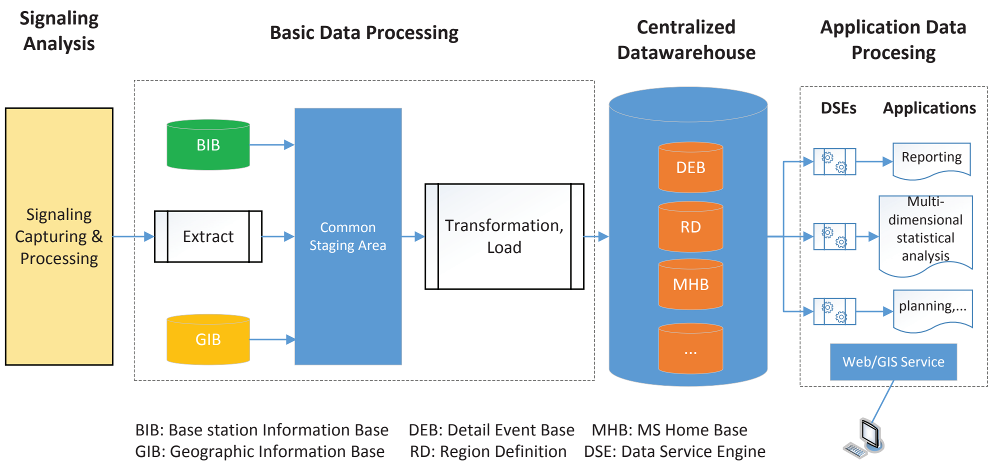


Fig. 1. Diagram of system architecture.

The basic data processing (BDP) part is to collect all sources of data and perform ETL (Extract, Transform and Load). Other than the CDRs from SA, there are two data sources: one is the base station information base (BIB) containing the position, LAC-CIs and other related information of each BS; the other is the geographic information base (GIB) containing the map layer data of the city. All data is integrated, transformed and loaded into the CDW by ETL tools.

The centralized data warehouse (CDW) part is the core of the system, containing all meta data, macro data and intermediate results. Three major data tables are the detail event database (DEB), region definition (RD) and MS home base (MHB). The DEB contains all recoded communication events appended with the defined region identifier. The RD contains all defined RoC with coordinates of their locating points, bordering roads, and etc. The MHB contains the mapping between the MSISDN prefix and MS registered province and cities.

The application data processing (ADP) part is to implement the data mining algorithms in *data service engines* (DESs), and present reports, multi-dimensional statistical analysis results to users. The system provides a Web/GIS interface for users to query analysis reports and define new regions.

B. Implementation

The system has been developed and deployed on the private cloud platform of the China Mobile Beijing Branch. Totally 16 unix/linux virtual machines (VMs) were utilized: 1 for interface server, 4 for Oracle DW clusters, 8 for analysis servers, 2 for GIS server and 1 for Web Server.

Signaling data mainly comes from 2G/3G network. Signaling messages from interfaces A and IuCS were captured, covering all 2G/3G subscribers of China Mobile Beijing branch. The interface A is the interface connecting BSC (Base Station Controller) and the MSC/MSS (Mobile Switching Center or MSC Server) in GSM networks, while interface IuCS is the interface between RNC (Radio Network Controller) and MSS in UMTS. The data is captured at the edge routers of MSC/MSSs. All signaling messages are parsed, correlated and transformed into communication event records or CDRs, containing the time stamps, the ciphered MSISDN, LAC, CI, event type ID, event info, number prefix and reserved fields. For privacy consideration and meeting anonymous monitoring requirement, the real MSISDN is ciphered using MD5 before entering into the system.

At present, total 128 regions have been defined and analysis reports are generated periodically. Aided by the GIS, administrator of the system can define a new region by drawing a closed area in the map. The real-time snapshot population, residence time, trajectory tracking and O-D analyses can be given immediately no later than 30–45 minutes dependent on different hours. Authorized users can access the website, browse the statistical tables and charts, print or download the analysis reports.

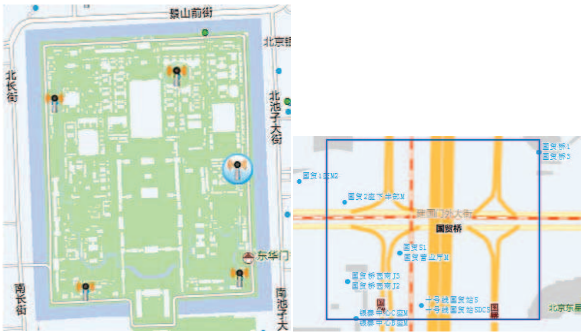
In this paper, the RSP, population home cities and their residence time characteristics are demonstrated.

V. Experiment and Results

A. The Observed Regions

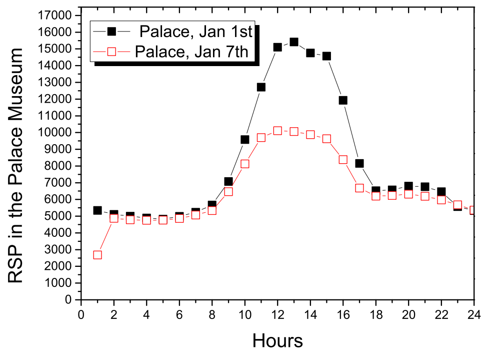
Two regions of concern were observed and compared: the Palace Museum, denoted by R_1, and the GuoMao Area, denoted by R_2, shown in Fig. 2a and Fig. 2b, respectively. R_1 is defined mainly for the Palace Museum, one of the most famous historical tourist attractions in Beijing; while R_2 is defined for an open area covering a 300m $\times$ 300m square, centering the GuoMaoQiao Overpass in the Beijing CBD.

Two regions have different features. R_1 is a typical closed ticketing area attracting tourists and its RSP is able to reflect the number variation of the tourists,

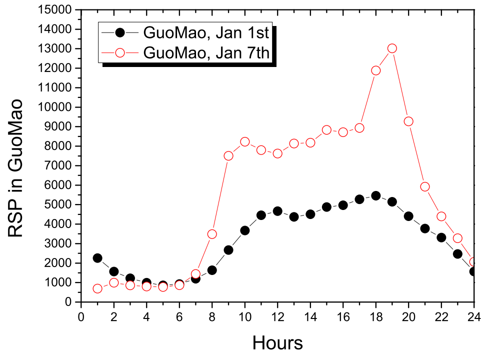


(a) R_1: the Palace Museum (b) R_2: the GuoMao Area

Fig. 2. Definitions the observed regions.



(a) R_1 on different days



(b) R_2 on different days

Fig. 3. Comparison of hourly RSP in different Regions and on different days.

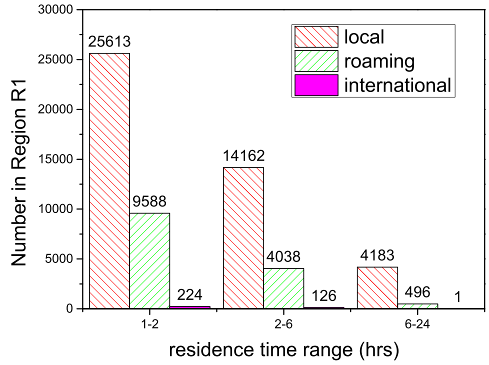
while R_2 is a public transport junction in the CBD and its RSP can reflect the variation of commuters to a great extent.

B. Hourly Snapshot Population

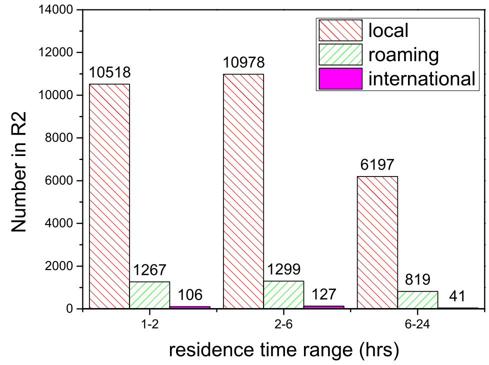
The hourly RSP in two regions was illustrated and compared in Fig. 3. In Fig. 3a, the hourly RSP in the region R_1 on a holiday and an ordinary working day was compared. As shown in Fig. 3a, their variation tendencies were very similar, but the absolute increase (the difference between the maximum and the ground) on the New Year Day was almost 2 times (1.97) that on the working day (Jan 7th, Tuesday).

In Fig. 3b, the hourly RSP in R_2 on different days was compared. Completely different from the R_1, traffic on a working day was extremely larger than on a holiday. On Jan 7th, Tuesday, it had a sharp increase from 5 pm and reached the maximum of 13021 at 7 pm, about 2.38 times the peak on the New Year Day.

The hourly RSPs in two regions on different days can be compared by combining Fig. 3a and 3b. On the New Year Day (Jan 1st) in R_1, traffic started to significantly rise from 8 am, and reached its maximum at 1 pm, then declined to almost the original value at 6 pm. In contrast, traffic in R_2, as a transport hub in CBD, was much less. It had a minimum in 5 am and slightly climbed up to its first peak at noon; then after a small decrease at 1 pm, it continued to increase gradually and reached its maximum in the day at 6 pm; and then went down slowly. On a working day, the peak RSP in R_2 exceeded that in R_1 on a working day, and its peak position came later in the evening.



(a) R_1, Jan 1st



(b) R_2, Jan 7th

Fig. 4. Population distribution of residence time ranges.

C. Residence Time Analysis

Using the algorithm described in Section III-B2, we can evaluate the characteristics of residence time for the population in different regions. The results for the regions R_1 and R_2 were demonstrated in Fig. 4, on different days, respectively.

The population was first divided into three groups: local, national roaming and international roaming, based on their country code and mobile number prefix. In Fig. 4a, population in R_1 on Jan 1st within different ranges of residence time was illustrated. As shown in the figure, a majority of the people spent 1–2 hours in the Palace Museum and less spent longer time. The population staying there for a longer time was only 51.7% and 13.2% of those visiting there within 1–2 hours. Also, the percentage of national roaming visitors decreased as the residence time increased, from 27.07%, 22.03%, down to 10.60%.

The situation was different in the Region R_2, illustrated in Fig. 4b. In the CBD center, the population with a residence time between 2–6 hours slightly exceeded those with a residence time of 1–2 hours by about 4.3%. The population staying within 6–24 hours was still 59.3% of those visiting there within 1–2 hours. It was consistent with the characteristics of a business area, significantly different from the distribution in the tourism region R_1.

D. Precision Consideration

In order to realize fast estimation and quasi real-time monitoring, we select cell-Id based positioning technology. In general, its positioning accuracy is dependent on the cell size. In urban areas, the cell size for 2G/3G mobile network is about $100 \sim 150$ metres. This accuracy is acceptable for the scenario of fast estimation, but still leaves room to improve.

The main inaccuracies originate from the following sources. First, mobile users belong to different network operators, therefore, MSs attaching to other network except China Mobile cannot be counted in. Second, MSs in the cells of BS near the regional border may not really locate within the RoC. Finally, the percentage of silent and power-off MSs affects the final precision. The movements of power-off ones cannot be detected by the network.

Nevertheless, it provides a practical approach to evaluate the population in a certain region, especially for an open area. Especially, in some cases, it is very helpful for the administration to identify the home of the population.

VI. Conclusion

In this paper, a monitoring system of snapshot population within a certain region was reported, based on cell-ID mobile positioning technology. The underlying schema for deriving the snapshot population and residence time through the time-spatial sequential datasets of discrete communication events were presented. The system architecture and data warehouse design were discussed; analysis results in two typical regions were demonstrated and thoroughly compared, including the hourly snapshot population and distributions of residence time. The system has been deployed and is monitoring the China Mobile subscribers in the whole city of Beijing. It has become a helpful tool serving several municipal departments and industries.

Due to the nature of cell-Id based positioning technology, the accuracy is not satisfactory. Other novel technologies like LTE (Long Term Evolution) networks based or indoor positioning technologies will be taken into consideration in the future work.

Acknowledgment

The work is jointly supported by the funding from the Chongqing Municipal Application and Development Planning Projects (cstc2013yykfA40006) and Program for Innovation Team Building at Institutions of Higher Education in Chongqing (KJTD201312). The authors would also like to thank the China Mobile Beijing Branch and Chongyou Huice Corporation for providing support of network data.

References

[1] Y. Zhao, "Mobile phone location determination and its impact on intelligent transportation systems," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 1, no. 1, pp. 55–64, 2000.

[2] J. Borkowski, J. Niemela, and J. Lempiainen, "Performance of Cell ID+RTT hybrid positioning method for UMTS radio networks," in *Proceedings of the 5th European Wireless Conference*, Barcelona, Spain, Feb 2004, pp. 487–492.

[3] D. Gundegard, A. Akram, S. Fowler, and H. Ahmad, "Cellular positioning using fingerprinting based on observed time differences," in *Smart Communications in Network Technologies (SaCoNeT), 2013 International Conference on*, vol. 01, 2013, pp. 1–5.

[4] D. Gundlegard and J. M. Karlsson, "Road traffic estimation using cellular network signal in intelligent transportation systems," in *Wireless technologies in Intelligent Transportation Systems*, Y. Z. Ming-Tuo Zhou and L. T. Yan, Eds. Nova Science Publishers, 2009.

[5] R. Bolla and F. Davoli, "Road traffic estimation from location tracking data in the mobile cellular network," in *Wireless Communications and Networking Confernce, 2000. WCNC. 2000 IEEE*, vol. 3, 2000, pp. 1107–1112 vol.3.

[6] S. Tao, V. Manolopoulos, S. Rodriguez, and A. Rusu, "Real-time urban traffic state estimation with A-GPS mobile phones as probes," *Journal of Transportation Technologies*, vol. 2, no. 1, pp. 22–31, Jan 2012.

[7] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '07. New York, NY, USA: ACM, 2007, pp. 330–339. [Online]. Available: http://doi.acm.org/10.1145/1281192.1281230

[8] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer, "Semantic trajectories: Mobility data computation and annotation," *ACM-TIST (ACM Transactions on Intelligent Systems and Technology)*, vol. 4, no. 3, pp. 49:1–49:38, 2012. [Online]. Available: http://infoscience.epfl.ch/record/177359

[9] F. Calabrese, G. D. Lorenzo, L. Liu, and C. Ratti, "Estimating origin-destination flows using mobile phone location data," *IEEE Pervasive Computing*, vol. 10, no. 4, pp. 36–44, 2011.

[10] N. Caceres, J. P. Wideberg, and F. G. Benitez, "Deriving origin destination data from a mobile phone network," *Intelligent Transport Systems, IET*, vol. 1, no. 1, pp. 15–26, 2007.

[11] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 4, pp. 1624–1639, 2011.

[12] I. A. Junglas and R. T. Watson, "Location-based services," *Commun. ACM*, vol. 51, no. 3, pp. 65–69, Mar. 2008. [Online]. Available: http://doi.acm.org/10.1145/1325555.1325568

[13] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-time urban monitoring using cell phones: A case study in rome," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, no. 1, pp. 141–151, 2011.