Steps towards the Extraction of Vehicular Mobility Patterns from 3G Signaling Data

Pierdomenico Fiadino[1], Danilo Valerio[1], Fabio Ricciato[1,3], and Karin Anna Hummel[2]

[1] Forschungszentrum Telekommunikation Wien (FTW), Austria
[2] ETH Zurich, Switzerland
[3] University of Salento, Italy

Abstract. The signaling traffic of a cellular network is rich of information related to the movement of devices across cell boundaries. Thus, passive monitoring of anonymized signaling traffic enables the observation of the devices' mobility patterns. This approach is intrinsically more powerful and accurate than previous studies based exclusively on Call Data Records as significantly more devices can be included for investigation, but it is also more challenging to implement due to a number of artifacts implicitly present in the network signaling. In this study we tackle the problem of estimating vehicular trajectories from 3G signaling traffic with particular focus on crucial elements of the data processing chain. The work is based on a sample set of anonymous traces from a large operational 3G network, including both the circuit-switched and packet-switched domains. We first investigate algorithms and procedures for preprocessing the raw dataset to make it suitable for mobility studies. Second, we present a preliminary analysis and characterization of the mobility signaling traffic. Finally, we present an algorithm for exploiting the refined data for road traffic monitoring, i.e., route detection. The work shows the potential of leveraging the 3G cellular network as a complementary "sensor" to existing solutions for road traffic monitoring.

1 Introduction

Road congestions cause safety dangers, unwanted CO_2 emissions, and major economic losses. Informing the drivers about real-time traffic status helps to reduce individual travel times, to optimize traffic flows, and to make a more efficient use of the road infrastructure. The systems currently used for gathering information about road traffic conditions are mainly based on road sensors, which would be very expensive to deploy on a large scale. Therefore, only some critical road sectors and junctions are covered by such monitoring infrastructure.

Cellular networks can help to overcome this problem: millions of mobile devices held by car drivers and passengers can be used opportunistically as road traffic probes, without facing the costs of deploying any new infrastructure. The main goal of this work is to develop a system that extracts road mobility data from the signaling traffic in a cellular network and uses the obtained information both for real-time applications and historical data analysis. Real-time applications include, e.g., inference of current road traffic intensity, detection of

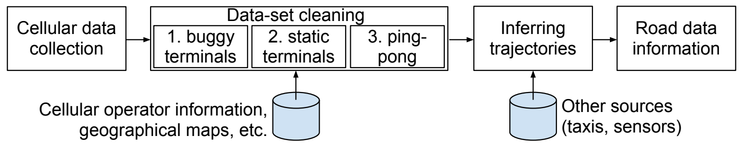


Fig. 1. Vehicular traffic monitoring approach including cellular data.

congestions, and estimation of expected travel times. Historical data analysis can provide support for tasks like road infrastructure planning, urban planning, optimization of public transport, etc.

The development of such a system must face several challenges. First, cellular networks produce an enormous amount of signaling data. Thus, efficient methods are needed to filter out data that are not relevant for the target task, e.g., from static or off-road terminals. Second, the mobility *perceived* by the network does not always map directly to the *actual* geographical mobility of the devices, due to the dynamics of Mobility Management (MM) protocols and radio propagation effects (e.g., fading). Third, cellular protocols are designed to minimize the signaling traffic through the network. Devices with no data or voice activity are tracked at a coarser spatial accuracy than active ones, i.e., at the level of location/routing area instead of cell. The system must be able to cope with this aspect, e.g., by reconstructing the state of the device by finite state machine.

Taking into account these challenges, we present a flexible traffic monitoring infrastructure on top of the cellular network that (a) collects signaling data from the network probes, (b) pre-processes the data and filters out irrelevant and/or spurious information, and (c) infers mobility trajectories across roads in order to pave the way towards the extraction of mobility patterns. In this paper, we describe the steps necessary to extract mobility patterns and include a desription of the artifacts emerged during the analysis of real-world cellular data.

The remainder of this paper is organized as follows: In Section 2, we describe the monitoring framework and present the main components of the cellular network architecture. In Section 3, we explore important examples of pre-processing filters that are required for reliably extracting mobility patterns. In Section 4, we present a first analysis and characterization of the dataset, and in Section 5 we introduce an algorithm for trajectory estimation. In Section 6, we relate our approach to other works. Finally, in Section 7 we draw the main conclusions.

2 Monitoring Framework

The analysis in this work was conducted on a dataset of anonymized signaling traffic captured during an entire working day in an Austrian cellular network. Figure 2 shows the main components of the monitoring system along with an overview of the cellular network architecture.

The cellular infrastructure is composed of a Core Network (CN) and a Radio Access Network (RAN). The CN is divided in two distinct domains: Circuit-Switched (CS) and Packet-Switched (PS). Mobile terminals can “attach” to the CS for voice call services, to the PS for packet data transfer, or to both. They can do so from 2G (GSM/EDGE) or 3G (UMTS/HSPA) radio bearers. Radio

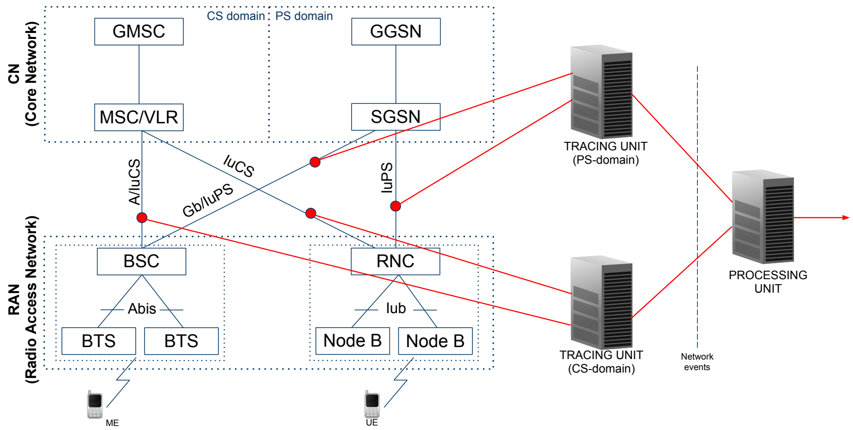


Fig. 2. Monitoring framework integrated into a cellular network architecture.

communication occurs between a mobile terminal and a base station serving a cell, which is the smallest spatial entity in the cellular network. Cells are grouped into larger logical entities: Routing Areas (RAs) and Location Areas (LAs) for the PS and CS domain respectively. In order to remain reachable, the terminals always inform the CN whenever they change LA and/or RA.

Our monitoring system collects signaling from the links between the cellular RAN and CN as indicated in Fig. 2, specifically on IuPS, IuCS, Gb and A interfaces. The amount and accuracy of mobility information that can be gathered from such interfaces varies with the terminal state: in general, the location of the terminals with an ongoing voice or data connection (active terminals) is known at the cell level, while the location of idle terminals is known only at LA/RA level. The network probes are connected to a couple of tracing units that interpret the signaling traffic and generate event-based tickets that are then forwarded to a processing unit. These tickets do not contain any identification of the devices nor payload of user traffic. In order to correlate and aggregate trajectories, each device is assigned an anonymous fingerprint generated by a one-way hash function of the corresponding International Mobile Subscriber Identity (IMSI) number. To further preserve user privacy, the hash key is regenerated daily with different seeds. For additional details about the monitoring system and the relevant MM protocols refer to [1, 2]. A detailed overview of the cellular network technology can be found in [3].

3 Mobility and Dataset Filtering

The ultimate goal of this work is to estimate vehicular trajectories based on *(i)* the passively monitored signaling traffic and *(ii)* the (known) geographical position and antenna orientation of the base stations. Hereby, defining a trajectory is not trivial: base station deployment parameters — e.g., directions and ranges of the antennas — and MM protocol dynamics must be taken into account to map cell sequences to a geographical trajectory. The raw dataset must undergo some preliminary filtering steps, namely:

(*1*) filtering traffic from buggy terminals and data inconsistencies;
(*2*) identifying and filtering traffic from static terminals;
(*3*) filtering network mobility artifacts.

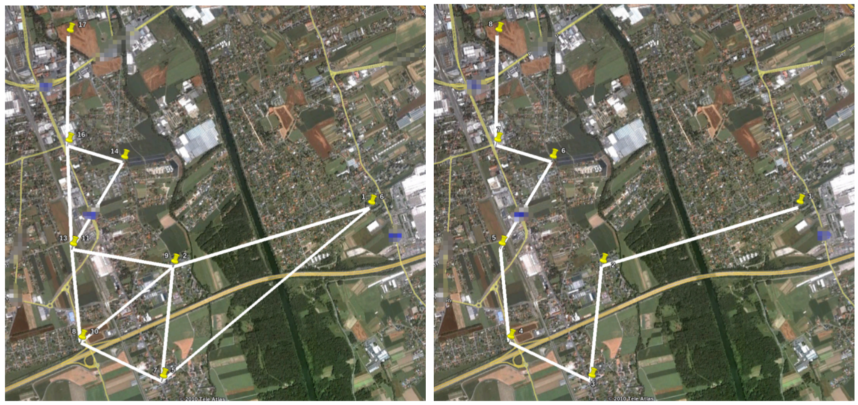
Given the complexity of the monitored network, one cannot assume that traces collected in real-world networks exhibit only expected signaling patterns: anomalies caused by buggy terminals, data holes, and other effects due to the limitations of the monitoring system cannot be avoided. Buggy behaviors can be caused by crashes or freezes in the operating system or in the baseband processor firmware of the terminals, causing the device to generate repeatedly the same message to the network. Such spurious traffic pollutes the dataset with sequences that interfere with the data processing algorithms. This traffic can be easily filtered out by simple thresholding, i.e., filtering all devices generating more than i events in a time interval t (e.g., per minute), as done in [4].

Cellular networks serve a large share of static terminals that do not change location, e.g., home computer with 3G modem. Considering the enormous amount of data to process, it is convenient to eliminate such devices from the dataset so as to relax the requirements in terms of processing resources, without any impact on the final output quality. Again, a simple thresholding approach is sufficient, filtering out devices that visit less than c cells in a time interval t.

The mobility *perceived* by the network signaling does not always map to the *real* geographical mobility of the devices. One of the main effects that must be considered is the presence of so-called "ping-pong" patterns, i.e., the repeated handover of a device between two or more cells caused by fluctuations of the signal strength level in time and/or space due to fading. Terminals exposed to the ping-pong effect are perceived by the system as highly mobile, but actually they can be geographically static. It is convenient to eliminate such cases to prevent distortions in the route classification algorithms presented later in Section 5.

For this preliminary study we used a simple algorithm for the automatic detection and filtering of ping-pong events. Given two cells A and B, we define a *ping-pong generated hop* when a device moves from cell A to B, and back, within a time interval t. The script keeps track of the last two cells crossed by each terminal. Assuming that a device generates a signaling message in the n-th cell, the system knows the $n-1$ and $n-2$ previous cell-ids. If cell n differs from cell $n-1$, the script checks whether cell n is equal to cell $n-2$. If so the timestamps of these two cells are checked: if $timestamp_n - timestamp_{n-2} < t$ the hop is considered to be part of a ping-pong sequence and therefore it is filtered. The algorithm can be easily extended to detect ping-pongs among more than two cells. Some care must be taken in choosing the right time window t. Too large values might cause false positives, hence discarding of useful data of local mobility. For this study, where the focus is on highway scenarios, we used a conservative setting of $t = 1$. In fact, due to the high speed of highway users, fading fluctuations seen by such devices are rapid — also fading at large-scale in space maps to fluctuations at small-scale in time — hence most cell transitions due to ping-pong have short durations. Instead, identifying the optimal setting of t for urban areas is more challenging — we leave this point for further study.

We tested the algorithm experimentally on the road. Figure 3 shows an example of how one of our test-drives is perceived by the network. The test car was



(a) before (b) after

Fig. 3. Visualization of the ping-pong effect before (left) and after filtering (right).

equipped with a 3G handset and a GPS device while driving on a highway exit ramp. The picture on the left depicts the cell sequence seen in the raw dataset. The sequence contains 17 hops occurred in 8 different cells. On the right side the same trajectory is depicted after filtering the ping-pong effect. About 50% of the hops were caused by ping-pong effects, which could be filtered by our algorithm leading to a visibly smoother path.

4 Signaling Traffic Characterization

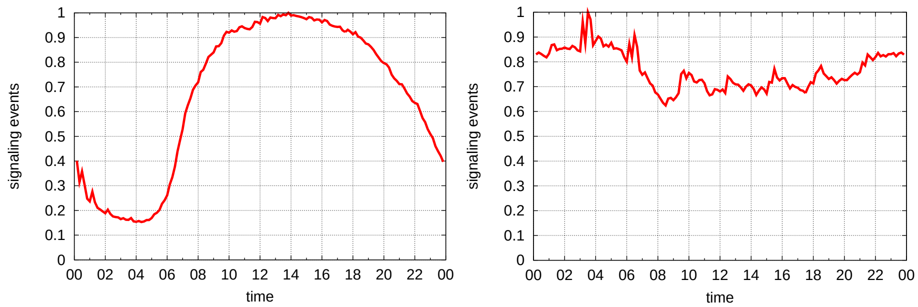
Given the complexity of the analysis task, it is convenient to familiarize with some basics aspects of the dataset at hand before delving into the algorithm details, e.g., looking at simple distributions of traffic across cells and/or terminals and exploring the principal different classes of terminal behaviour. In this section, we present some basic aspects of the MM signaling present in our sample dataset from a real-world cellular network in Austria. Where applicable, we show the effects of applying the filtering steps discussed in the previous section.

4.1 Time-charts of signaling events

Figure 4 shows the time series of the number of signaling events during an entire day, aggregated per minute[4]. The chart on the left shows the total amount of signaling events like cell handovers, LA updates, etc. The typical time-of-day pattern associated to the daily cycle of human behaviour is clearly visible.

On the right side we plot the time series of the *periodic LA updates*, i.e., a sort of "keep-alive" message that is sent by terminals that do not move nor engage in any active voice/data call for a certain time. Therefore, such traffic is anti-correlated with user mobility — the less they move, the more likely they produce periodic LA updates — and in fact the trend is somewhat opposite to the previous graph.

[4] In this and the following plots the values in the y-axis are normalized to prevent disclosure of data considered business-sensitive by the network operator.



(a) Global signaling (b) Periodic updates

Fig. 4. Signaling traffic observed in one day.

4.2 Number of crossed cells

The solid line in Figure 5 shows the number of unique cells visited in one day by each terminal, considering all signaling messages in the dataset. The plot shows that more than 15% of terminals appear only in a single cell, while approximately 50% generate signaling in less than five cells. This confirms the opportunity of introducing a preprocessing stage to remove static users from the raw dataset.

The dashed line emulates what would be seen by CDR: it reports the number of different cells where the user started or terminated a voice/data connection. More than 25% of the terminals generate/receive calls in a single cell during a whole day. The remarkable gap between the two curves shows that relying on CDRs produces a biased perception of the mobility of the population, leading to underestimate the actual level of user mobility. In other words, we conjecture that the *quantitative metrics* reported in CDR-based studies like, e.g., [5, 6], are not representative of the actual mobility process. A more accurate comparison between the mobility measurements obtained by CDR and CN/RAN signaling is part of our ongoing work.

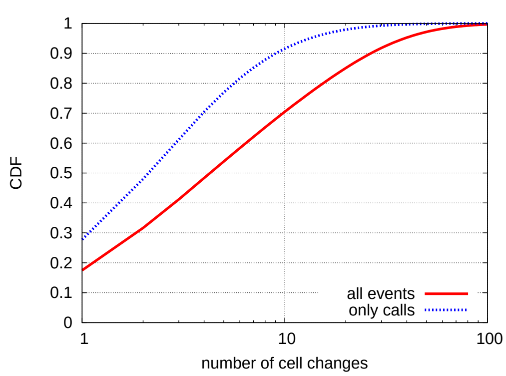
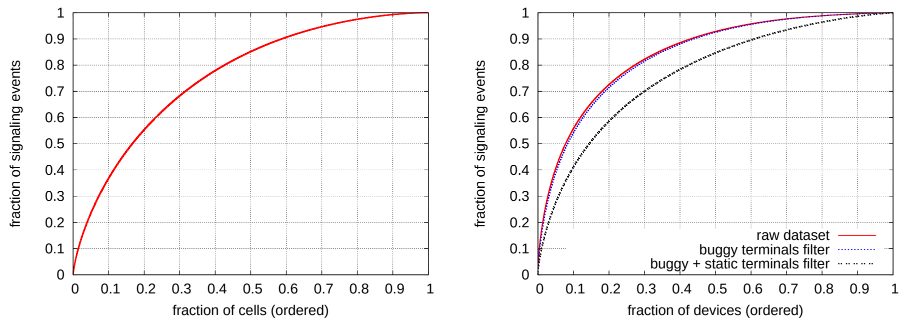


Fig. 5. Number of daily crossed cells



(a) versus cells

(b) versus devices

Fig. 6. Cumulative distribution of signaling traffic.

4.3 Signaling distribution across cells and subscribers

Figure 6(a) shows the distribution of the signaling traffic across cells. The plot reveals the existence of a small set of high-activity cells: 30% of the base stations are involved in the exchange of 70% of the signaling traffic. Analogously, Figure 6(b) depicts the distribution across terminals: considering the raw dataset, 20% of the devices attached to the network generate 75% of the total amount of signaling traffic. These types of distributions with large disparity are often found in cellular networks [7–12]. The question arises whether this distribution is representative for the signaling traffic even after filtering out buggy terminals and static users. The plot shows that the removal of buggy terminal does not have a remarkable effect on the distribution, thus the curves are almost overlapping. In fact, while these terminals have deleterious effects in the processing algorithms, they are too sparse to influence the total distribution. The same cannot be said for static users. When our filtering algorithm is applied (dotted line) the curve changes considerably, showing a more uniform distribution.

4.4 Idle vs. active devices

The previous section highlighted the existence of a large set of static mobile devices and a small set of highly mobile ones. Figure 7 shows the distribution of *cell update* signaling messages among devices. Such messages are produced only by user in active state, engaged in data and/or voice connection. It can be seen that there is a small class of terminals that generate a very large number of cell updates. These terminals are always active and provide the network with fine-grained information about their mobility.

An efficient system for road traffic flow analysis must consider mobility events from both idle and active users. At any time the vast majority of the terminals are in idle state. These provide a more complete view of the mobility flows than the fewer active terminals, but at a coarser spatial granularity (LA or RA level). Conversely, active terminals provide detailed information about their movements at cell level and can be used as single, sparse, and more accurate road traffic probes in addition to passive terminals. In some sense the two classes play complementary roles in terms of sample coverage vs. spatial accuracy. Thus,

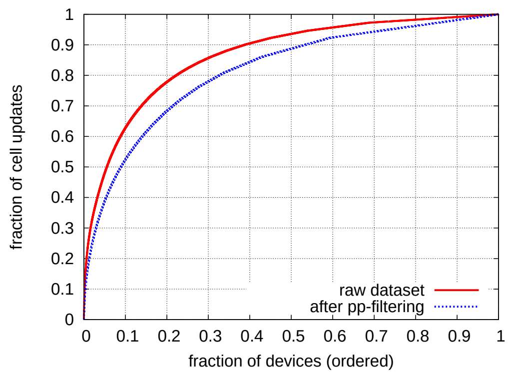


Fig. 7. Distribution of *cell updates* among devices.

in our approach we exploit data from both active and idle users to develop a system that dynamically reveals traffic jams, road anomalies, and travel times more accurately than alternative approaches based only on signaling traffic from active users such as [13, 14].

5 Route Detection

Route detection is a prerequisite for several road traffic studies based on cellular network signaling. The goal of route detection is to map handset transitions (handovers) among network entities to driver trajectories in the road network.

In this section, we propose a novel algorithm that is based on the *vector space classification* and takes inspiration from the field of Web information retrieval and relevance ranking. The key idea is the following: the sequence of base stations covering a path (road) and the sequence of cells crossed by terminals can be represented as vectors in a vector space. In this context, the problem of deciding whether a cellular user is traveling along a road is solved with the computation of the *similarity* of the respective user and road vectors: the higher the similarity, the higher the probability that the cellular user traveled along the corresponding road. The procedure for performing route detection is composed mainly by two steps: *vector generation* and *similarity computation*.

5.1 Vectors generation

We define a *road descriptor* as a static sequence of cells that cover the road in a specific direction of interest. In other words, the road descriptor represents the sequence of cells which a device is most likely to communicate with, when traveling the road in one direction. Road descriptors can be built by using information available in radio network planning departments of mobile operators or by simple test-drives aimed at building cell-handovers logs. Note that the sequence of cells differs depending whether the analysis focuses on idle or active devices, i.e., in case of idle devices it includes only cells at the LA borders.

We define a *user path descriptor* as the sequence of cells traversed by a terminal. This sequence starts being computed as soon as the device attaches to a cell

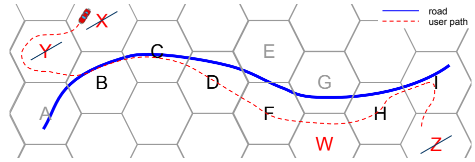


Fig. 8. Example of a user's cell sequence.

belonging to a road descriptor. Figure 8 shows an example: the road descriptor is (A,B,C,D,E,F,G,H,I), the terminal covers cells (X,Y,B,C,D,F,W,H,I,Z) and the resulting user path descriptor is (B,C,D,F,W,H,I).

The two descriptors are used to generate road and user vectors in the n-*dimensional* space, where n is the total number of unique cells in road and/or user sequence. Both vectors are binary and of the same dimension. The generation of user path descriptors and the procedure for building road and user vectors are formally described by pseudo-code in Algorithm 1 and Algorithm 2. Figure 9 shows an example of road and user vectors, where the road sequence is composed of 5 cells and the user traverses 2 additional cells (i.e., $n = 7$).

5.2 Similarity computation

The second step involves the computation of the similarity between user and road vectors. The similarity between two vectors is defined as the inverse of their geometrical distance and ranges between 0 and 1.

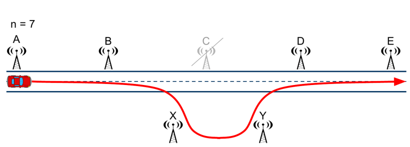
The similarity can be computed using several metrics. Some examples are the cosine similarity, the Jaccard index, and the Tanimoto coefficient [15]. The latter is the most appropriate for our aims, as it is designed for binary vectors and penalizes the presence of *external cells* more than other similarity metrics. In other words, it mitigates better the problem of ambiguous user placement in case of nearly parallel roads. Tanimoto similarity is defined as

$$T(A,B) = \frac{A \cdot B}{\|A\|^2 + \|B\|^2 - A \cdot B}$$

where A and B are two binary vectors and $A \cdot B = \sum_i (A_i \bigwedge B_i)$, $\|A\|^2 = \sum_i (A_i)$. We computed the Tanimoto similarity with the following equation:

$$similarity = \frac{(\#\, internal\ cells)}{(length\ of\ road\ descr) + (length\ of\ user\ seq) - (\#\, internal\ cells)}$$

thus, without the vector generation procedure but directly using road and user sequences. This allows us to release processing power for other tasks. Note that all previous considerations about vectorial route detection remain valid even when this simplified formula is applied.



	A	B	C	X	Y	D	E
Road vector	1	1	1	0	0	1	1
User vector	1	1	0	1	1	1	1

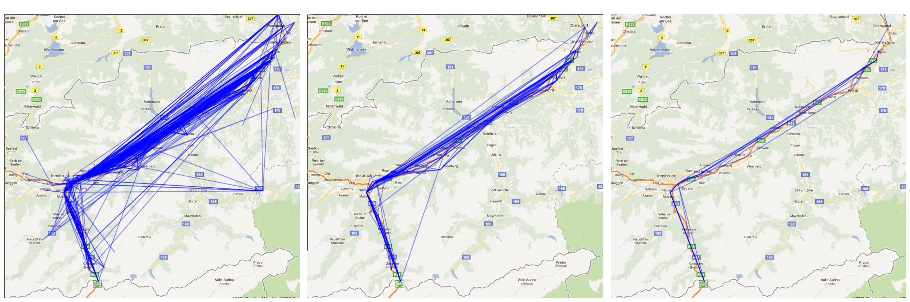
Fig. 9. Example of vectors generation.

Algorithm 1 UserSequenceExtraction (*user_sequence u*, *road_sequence r*)

```
k ← 0
start_flag ← 0
for i = 1 ... |u| do
    if u[i] ∈ r then
        start_flag ← 1
        k ← position of u[i] in r
        if j > k then
            k ← j
            append aux_vector to result_sequence_(u,r)
            append u[i] to result_sequence_(u,r)
            empty aux_sequence
        end if
    else
        if start_flag = 1 then
            append u[i] to aux_sequence
        end if
    end if
end for
return result_sequence_(u,r)
```

5.3 Route classification

Once all similarities have been computed, each user is assigned to the road whose vector presents higher similarity. This implies that there exists a road descriptor for each road in the region of interest. In reality, extracting road descriptors is a tedious procedure that requires a number of test-drives and detailed knowledge of the operator radio planning. Thus, it is not always possible to derive road descriptors for the entire road network. To overcome this problem we propose a threshold-based metric, i.e., a road sector is considered to be traveled by a terminal when the similarity between the user and the road vector is higher than a specific threshold. In the experiments we have performed so far, we manually tuned this parameter and analyzed the type and number of output trajectories. Validating such results is difficult due to the lack of ground truth to compare with. Note that according to the vector generation procedure described above,



(a) threshold = 0.2 (b) threshold = 0.4 (c) threshold = 0.6

Fig. 10. Comparison of different similarity thresholds.

Algorithm 2 VectorsGenerator ($filtered\_user\_sequence\ u$, $road\_sequence\ r$)

```
i ← 0
j ← 0
while i < |r| do
    if r[i] ≠ u[j] then
        while u[j] ∉ r do
            append 0 to road_vector
            append 1 to user_vector
            j++
        end while
        if u[j] = r[i] then
            append 1 to user_vector
        else
            append 0 to user_vector
        end if
    else
        append 1 to user_vector
        j++
    end if
    append 1 to road_vector
    i++
end while
return user_vector, road_vector
```

the similarity is computed for every pair of user-road vectors that share at least one common cell in the descriptors. Figure 10 shows an illustrative example of the effect of different threshold values. We compute the similarity of every mobile users' vector with a single sample road vector representing part of an Austrian motorway. It can be noted that the lower the threshold, the higher the amount of vehicles associated with the road segment. A too low value increases the probability of including users who didn't actually driven that road. A too high threshold reduces considerably the number of output trajectories, but potentially excludes some of the target users. The optimal value of the threshold depends on several factors, including the accuracy of the road vector, the radio network design, the existence of overlapping cells and umbrella cells, the length of the considered road segment, etc. Automatic tuning of this parameter is part of our ongoing work.

6 Related Work

The use of cellular networks for extracting vehicular mobility data has been the focus of several commercial products in the last years, aiming at selling cheaper and more scalable ways to infer and redistribute road traffic information [16–18]. Unfortunately, given their commercial nature, no public information is available describing algorithms and procedure used by these products, preventing the scientific community from analyzing and evaluating such systems.

A few reserch works and projects have been devoted to the topic and only a subset of them utilizes real-world data from operational mobile networks. In general, one can distinguish between active, passive, and application-based mobility monitoring. The most common systems rely on the application-based approach, i.e., smartphone applications that report GPS and cellular information to a central server. In [19, 20] the authors propose an application-based system for road traffic estimation using a small set of sample users. Sricharan et al. [21] use the same approach for the characterization of users' mobility and clustering in homogeneous groups. Analyzing a large amount of user reported GPS information is the aim of other projects such as the Mobile Millennium Project [22]. GPS data is particularly suitable for studying micro-mobility characteristics, including velocity, directional changes, and start time of trips, as shown in our previous work [23], but data sets are usually smaller than those from cellular networks.

Most of previous studies based on passive data from the cellular network make use of (anonymized) Call Data Records (CDRs) [5–9]. CDR are summary tickets produced for billing purposes at every voice call, data connection or SMS envoy. With CDR the user location can be observed only at the time of initiating or terminating a call, data connection or SMS. In general, this approach gives a limited and somewhat biased view of user mobility, as his trajectory is sampled at instants that are dependent on (conditioned to) his calling activity, which in turn might depend on factors like time-of-day, position etc. The limitation of the CDR approach is particularly serious when considering road users, that in general are less likely to engage in calls during the trip.

Trestian *et al.* [13, 14] adopted a passive monitoring approach that is somehow intermediate between the CDR method and the one we adopt in this work. By monitoring the CN links in the PS domain they are able to observe the user location only during data connections, i.e., for active terminals. Instead, by monitoring the links between the CN and the RAN, we collect mobility information also for terminals in idle state. Furthermore, our dataset covers both the PS and CS domains, and both 2G (GSM/EDGE) and 3G (UMTS/HSPA) radio access.

Two other projects, namely *Traffic.online* [24] and *Do-iT* [25], consider passive monitoring data from the RAN, and specifically the A-bis interface. This approach provides very accurate information but is also expensive, as it might require to tap hundreds or even thousands of A-bis links to cover a nation-wide network, depending on the particular network deployment.

Besides empirical analyses of real-world data, the literature is rich of studies based on network-level simulations. Among them, our work is closest related to the work of Gundlegard et al. [26, 27] who state goals similar to ours but apply different methodologies. A detailed survey of the state-of-the-art in road traffic estimation from cellular mobility signaling can be further found in one of our previous publications [1].

7 Conclusions

In this paper, we presented a work-in-progress study on the potential of using signaling traffic from the mobile network for road mobility analysis. The work is based on a real-world dataset from an Austrian cellular network operator.

We showed the main processing steps that must be taken to extract road mobility patterns from the signaling traffic stream, highlighting the necessary pre-filtering procedures. We showed that the signaling dataset contains a low amount of highly mobile users. We also discussed the importance of considering both idle and active terminals, that provide complementary views in terms of coverage and spatial granularity. Finally, we presented a method to map "road descriptions" and "user paths", both in terms of cell sequences, by applying a distance metric for binary vectors.

As for our future work, we are currently collecting data from a set of reliable alternative sources and building an accurate ground-truth for the validation of the presented results. In particular, we are conducting several test-drives in many well-traveled Austrian motorways. In each run we log all events reported by the radio interface layer of GPS-equipped smartphones and correlate the actual mobility (i.e. GPS based) with the mobility perceived by our system in terms of traveled road sections. In addition, we are performing comparisons between the traffic intensities reported by selected road sensors of the motorway operator and the ones perceived by our system. Preliminary results are very encouraging. Finally, we are tuning the parameters of the algorithms presented throughout this paper, with particular focus on investigating and identifying the optimal threshold values for the ping-pong filter and the Tanimoto similarity. Our final aim is to design a road monitoring module that continuously analyzes cellular network signaling and infers current traffic conditions reliably and *in real-time*. We envision a system that is able to detect anomalies using particular events in the signaling traffic — e.g., a drop in the handover rate or sudden change in the number of road user across a road segment — and combine current road condition with past history in order to predict upcoming road congestions, i.e., traffic forecast.

In another parallel work we aim at comparing the different views that the three approaches to mobility estimation based on cellular network data — namely CDR [5, 6], CN/PS [13, 14] and our approach — can provide about the underlying human mobility patterns.

Acknowledgment

This work has been supported by the Austrian Government and by the City of Vienna within the competence center program COMET. The work of K.A. Hummel has been supported by the Commission of the European Union under the FP7 Marie Curie IEF program contract PIEF-GA-2010-276336 MOVE-R.

References

1. D. Valerio et al. Road traffic estimation from cellular network monitoring: a hands-on investigation. In *IEEE Personal Indoor Mobile Radio Communication Simposium 2009 (IEEE PIMRC '09)*, September 2009.
2. D. Valerio et al. Exploiting cellular networks for road traffic estimation: a survey and a research roadmap. In *IEEE VTC'09-Spring*, April 2009.
3. H. Holma and A. Toskala. *WCDMA for UMTS : Radio Access for Third Generation Mobile Communications*. John Wiley and Sons, Ltd., 2000.

4. F. Ricciato et al. Measurement-based optimization of a 3g core network: a case study. In *6th Int'l Conference on Next Generation Teletraffic and Wired/Wireless Advanced (NEW2AN 2006)*, St. Petersburg, Russia, June 2006.
5. M. C. González, C. A. Hidalgo, and A. L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(5), June 2008.
6. C. Song, Z. Qu, N. Blumm, and A. L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968), February 2010.
7. E. Halepovic and C. Williamson. Characterizing and Modeling User Mobility in a Cellular Data Network. In *2nd ACM Int'l Workshop on Performance Evaluation of Wireless Ad Hoc, Sensor, and Ubiquitous Networks (PE-WASUN)*, October 2005.
8. S. Isaacman et al. A Tale of Two Cities. In *11th Workshop on Mobile Computing Systems and Applications (HotMobile)*, Annapolis, Maryland, February 2010.
9. S. Isaacman et al. Ranges of Human Mobility in Los Angeles and New York. In *8th Int'l Workshop on Managing Ubiquitous Comm. and Services*, March 2011.
10. A. Coluccia et al. On robust estimation of network-wide packet loss in 3g cellular networks. In *5th IEEE Broadband Wireless Access Workshop (BWA'09)*, Dec 2009.
11. A. Coluccia et al. Bayesian estimation of network-wide mean failure probability in 3g cellular networks. In *PERFORM 2010 Workshop*, October 2010.
12. A. D'Alconzo et al. A distribution-based approach to anomaly detection for 3g mobile networks. In *IEEE GLOBECOM*, December 2009.
13. I. Trestian et al. Measuring serendipity: connecting people, locations and interests in a mobile 3g network. In *Internet Measurement Conference (IMC'09)*, Chicago, IL, November 2009.
14. I. Trestian et al. Taming user-generated content in mobile networks via drop zones. In *IEEE INFOCOM 2011*, Shanghai, China, April 2011.
15. Tanimoto. Internal report. Technical report, IBM, Nov. 1957.
16. TomTom Traffic, http://www.tomtom.com/landing_pages/traffic_solutions/.
17. Cellint Traffic Solutions http://www.cellint.com.
18. Decell Technologies Ltd. http://www.decell.com.
19. A. Poolsawat et al. Acquiring road traffic information through mobile phones. In *8th Int'l Conference on ITS Telecommunications*, October 2008.
20. Quang Tran Minh and E. Kamioka. Granular quantifying traffic states using mobile probes. In *IEEE VTC 2010-Fall*, October 2010.
21. M.S. Sricharan and V. Vaidehi. A Pragmatic Analysis of User Mobility Patterns in Macrocellular Wireless Networks. In *World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, Espoo, Finland, June 2007.
22. Mobile Millennium, Project http://traffic.berkeley.edu.
23. K.A. Hummel and A. Hess. Estimating Human Movement Activities for Opportunistic Networking: A Study of Movement Features. In *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2011)*, pages 1–7, Lucca, Italy, June 2011.
24. Traffic.online, http://www.its-munich.de/pdf/DynInfoBY/4-Birle.pdf.
25. Do-iT T-Mobile D1.
26. D. Gundlegard and J. M. Karlsson. Route Classification in Travel Time Estimation Based on Cellular Network Signaling. In *12th Int'l IEEE Conference on ITS*, 2009.
27. D. Gundlegard and J. M. Karlsson. Handover location accuracy for travel time estimation in GSM and UMTS. In *IET Intelligent Transport System*, March 2008.