Towards Real-Time Road Traffic Analytics using Telco Big Data

Constantinos Costa, Georgios Chatzimilioudis
University of Cyprus
{costa.c,gchatzim}@cs.ucy.ac.cy

Demetrios Zeinalipour-Yazti
Max Planck Institute for Informatics and University of Cyprus
dzeinali@mpi-inf.mpg.de

Mohamed F. Mokbel
University of Minnesota
mokbel@cs.umn.edu

ABSTRACT

A telecommunication company (telco) is traditionally only perceived as the entity that provides telecommunication services, such as telephony and data communication access to users. However, the IP backbone infrastructure of such entities spanning densely urban spaces and widely rural areas, provides nowadays a unique opportunity to collect immense amounts of mobility data that can provide valuable insights for road traffic management and avoidance. In this paper we outline the components of the *Traffic-TBD (Traffic Telco Big Data)* architecture, which aims to become an innovative road traffic analytic and prediction system with the following desiderata: i) provide micro-level traffic modeling and prediction that goes beyond the current state provided by Internet-based navigation enterprises utilizing crowdsourcing; ii) retain the location privacy boundaries of users inside their mobile network operator, to avoid the risks of exposing location data to third-party mobile applications; and iii) be available with minimal costs and using existing infrastructure (i.e., cell towers and TBD data streams are readily available inside a telco). Road traffic understanding, management and analytics can minimize the number of road accidents, optimize fuel and energy consumption, avoid unexpected delays, contribute to a macroscopic spatio-temporal understanding of traffic in cities but also to "smart" societies through applications in city planning, public transportation, logistics and fleet management for enterprises, startups and governmental bodies.

CCS CONCEPTS

• **Information systems → Database management system engines**; **Online analytical processing engines**; **Spatial-temporal systems**; **Geographic information systems**;

KEYWORDS

Telco, Big Data, Road Traffic, Data Analytics

ACM Reference Format:
Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, and Mohamed F. Mokbel. 2017. Towards Real-Time Road Traffic Analytics using Telco Big Data . In *Proceedings of BIRTE'17, Munich, Germany, August 28, 2017,* 5 pages.
https://doi.org/10.1145/3129292.3129296

BIRTE'17, August 28, 2017, Munich, Germany

ACM ISBN 978-1-4503-5425-7/17/08…$15.00
https://doi.org/10.1145/3129292.3129296

1 INTRODUCTION

Real-time road traffic is an open problem in cities that target the avoidance of road accidents, fuel and energy consumption and unexpected delays. According to INRIX [10], a global leader in transportation analytics, the U.S. was ranked as the most traffic congest country across 38 countries in 2016. The INRIX traffic scorecard shows that the congestion costs U.S. drivers nearly $300 Billion USD in 2016, which is an average of $1400 per driver per year (both direct and indirect costs). At a city level, Los Angeles (depicted in Figure 1, left), ranked as the most congested city with an average of 104 hours in traffic jams during peak congestion. This resulted in a total cost of $9.6 Billion USD with an average cost of $2480 per driver. Providing flexible and high resolution traffic analytics, beyond aggregate statistics, can enable a wide range of enterprise, startup and governmental bodies to develop next generation decision support systems necessary for managing and planning the ever more complex future urban spaces.

A primitive component for high-resolution traffic analytics is the task of accurate *traffic mapping*, i.e., identifying the population of cars on segments of a road network. First Era (1E) traffic mapping methods, as we coin them in this work, have relied on dedicated hardware, such as magnetic loop detectors or traffic cameras mounted on traffic lamps or road-sides [1]. The problem with 1E systems was that these have been too invasive, expensive to install and maintain but also could not adequately scale to large urban and rural areas. The advent of smartphones gave birth to *Second Era (2E)* traffic mapping systems that were handheld-based. In 2E systems, a variety of "probe vehicles", such as Taxis [21] or Buses [25], were equipped with a smartphone and a dedicated app that would supplement a service with location updates on an ongoing basis. Successful commercial 2E systems are Google Maps (e.g., through their Android OS) or Here Maps (i.e., Nokia/Microsoft OS now owned by Audi, BMW and Mercedes). A somehow improved approach to 2E systems were what we coin the 2.5E systems, which would use ordinary car drivers for contributing data in both an *opportunistic* crowdsourcing manner (e.g., a user allows its device to report its location to the service) but also a *participatory* crowdsourcing manner (e.g., a user reports on a map where a road was blocked). Successful 2.5E systems include Google Maps, Apple Maps, Waze, Nokia's HERE maps, and Mapquest.

The problem with state-of-the-art 2E/2.5E systems are as follows: i) these rely on the full cooperation of probe vehicle that is many times difficult [25]; ii) these expose the location of users to third-party mobile applications, which introduces privacy concerns; and iii) these don't provide a fine-grain mapping of traffic especially in complex urban areas (e.g., arbitrary road segments are typically shown to have traffic on popular navigation systems

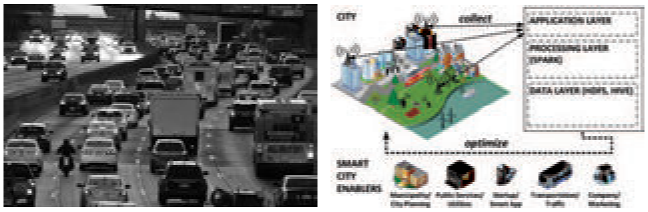

Figure 1: (Left) Traffic Analytics are critical for improving the quality of life for millions of users across the world. Los Angeles traffic had a total cost of $9.6 Billion USD, i.e., an average cost of $2480 per driver. (Right) The proposed *Traffic Telco Big Data (Traffic-TBD)* architecture aims to contribute to a macroscopic spatio-temporal understanding of traffic but also to "smart" societies through applications in city planning, public transportation, logistics and fleet management for enterprises, startups and governmental bodies.

even though the traffic has already been dissolved or was never there).

In recent years there has been considerable interest from *telecommunication companies (telcos)* to extract concealed value from their network data. Consider for example a telco in the city of Shenzhen, China, which serves 10 million users. Such a telco is shown to produce 5TB per day [23] (i.e., thousands to millions of records every second). Huang et al. [9] break their 2.26TB per day *Telco Big Data (TBD)* down as follows: (i) *Business Supporting Systems (BSS)* data, which is generated by the internal work-flows of a telco (e.g., billing, support), accounting to a moderate of 24GB per day and; (ii) *Operation Supporting Systems (OSS)* data, which is generated by the Radio and Core equipment of a telco, accounting to 2.2TB per day and occupying over 97% of the total volume. Particularly, the OSS.MR data [4], i.e., signaling data from cell towers, can be used to triangulate user trajectories providing localization accuracies that are typically below 50 meters [27].

In this position paper we outline the components of the *Traffic-TBD (Traffic Telco Big Data)* architecture, which aims to become an innovative road traffic analytic and prediction system with the following desiderata: i) provide micro-level traffic modeling and prediction that goes beyond the current state provided by Internet-based navigation enterprises that utilize crowdsourcing (i.e., 2.5E systems); ii) retain the location privacy boundaries of users inside their mobile network operator, to avoid the risks of exposing location data to third-party mobile applications; and iii) be available with minimal costs and using existing infrastructure (i.e., cell towers and TBD data streams readily available in the telco architecture). Traffic-TBD is what we call a 3E system, which relies both on data generated by 2.5E systems and TBD data to generate the traffic mapping both accurately, without additional hardware and additional costs.

The Traffic-TBD architecture is envisioned on top of our SPATE architecture [4] and consist of the following conceptual layers: the *data layer*, which is responsible to aggregate the incoming and stored data in an online and offline mode and provide temporal decaying and compressed access methods to various input sources (e.g., telco data, Map data, even social data); the *processing layer*, which is providing the mechanisms for minimizing the query response time for data analytic queries; and iii) the *application layer*, which consists of a variety of components to support a range of domain-specific analytic tasks we will outline in this paper.

In the industrial sphere, the most similar effort is the TomTom-HD Traffic system [3], which obtains 3E traffic mapping from TomTom-HD traffic users across the Vodafone network as they move around the road networks. TomTom-HD has the following high-level differences from Traffic-TBD: i) it only uses a spatial visual analytic interface, as opposed to spatio-temporal interface proposed by Traffic-TBD. Effectively, it lacks the visual cues to study traffic at various time instances (e.g., using heatmaps, motion heatmaps and videos); ii) it focuses more on plan routing functionality as opposed to spatio-temporal traffic analytics (i.e., traffic flows and traffic incidents, clustering and classification, learning and prediction all presented in Section 3); iii) its a closed system, so we can only speculate about its internal architecture.

2 BACKGROUND AND RELATED WORK

In this section we start out with a general overview of telco big data research. We then focus on the specific topic of traffic analytics, which lies at the epicenter of this work.

2.1 Telco Big Data (TBD) Research

TBD research falls in the following categories: (i) real-time analytics and detection; (ii) experience, behavior and retention analytics; and (iii) privacy. There is also traditional telco research not related to big data, rather comprises of topics related to business (BSS) data in relational databases. Zhang et al. [23] developed *OceanRT*, which was one of the first real-time telco big data analytic demonstrations. Iyer et al. [11] present *CellIQ* to optimize queries such as *"spatiotemporal traffic hotspots"* and *"hand-off sequences with performance problems"*. It represents the snapshots of cellular network data as graphs and leverages on the spatial and temporal locality of cellular network data. Zhu et al. [27] deal with the usage of telco MR data for city-scale localization, which is of particular interest in the mapping of the Traffic-TBD architecture. Huang et al. [9] empirically demonstrate that customer churn prediction performance can be significantly improved with telco big data. Luo et al. [16] propose a framework to predict user behavior involving more than one million telco users. Ho et. al. [6] propose a distributed community detection algorithm that aims to discover groups of users that share similar edge properties reflecting customer behavior. Hu et al. [8] study Differential Privacy for data mining applications over telco big data and show that for real-word industrial data mining systems the strong privacy guarantees given by differential privacy are traded with a 15% to 30% loss of accuracy. Privacy and confidentiality are critical for telcos' reliability due to the highly sensitive attributes of user data located in CDR, such as billing records, calling numbers, call duration, data sessions, and trajectory information.

2.2 Traffic Analytics

Real-time Analytics and Detection: Real-time road traffic is an open problem in cities that target the avoidance of road accidents, fuel and energy consumption and unexpected delays. 1E solutions

Figure 2: (Left) The components of the proposed Traffic-TBD architecture. (Right) A preliminary version of the spatio-temporal Traffic-TBD analytic interface.

have been using expensive hardware sensor data [2]. As described in the introduction, there are 2E and 2.5E based on floating vehicle data [15, 21] and crowdsourced data [7, 12] including data from social networks [20]. Pham et al. [18] developed a system for detection of road traffic in real-time using mobile devices. The system was using low and high frequency motion detection and activity classification with precision and recall over 83% for traffic detection. A huge investment should be made in order to provide data using hardware infrastructure like cameras. In contrast, crowdsourced data could provide valuable insights in order to estimate the road traffic. In particular, Jain et al. [12] implemented a MapReduce system that generates road traffic analytics efficiently. In addition, Hu et al. [7] presented SmartRoad, which is a crowdsourced system the detects traffic regulators, traffic lights and stop signs using the sensing capabilities of modern smartphones and the power state of the car. Janecek at al. [13] studied how the cellular networks could be used in order to determine the traffic data in real time minimizing the social, environmental and financial constraints for the necessary hardware to monitoring the road network. This will support the real-time road traffic monitoring and enable the road traffic administration office to act efficiently. Privacy is still an open problem due to sensitive spatio-temporal information. Goel et al. [5] introduce k-anonymous and l-grouping for trajectory estimation providing a trade-off between accuracy and privacy.

Traffic Prediction: Predicting road traffic is an important functionality that can facilitate route planning and other traffic decision making processes. Zheng et al. [24] proposed a system that uses inexpensive big data collected from buildings information to predict the traffic in the proximate area. Zhou et al. [26] incorporated data from historical road monitoring systems and online information about events and weather achieving an overall accuracy of 93%. Xu et al. [22] study how the predicted traffic models based on historical data are different from the real-time traffic situations. The authors proposed a novel framework that is using data in real-time for traffic prediction by matching the current traffic context with the most suitable prediction model based on the historical data.

3 THE TRAFFIC-TBD ARCHITECTURE

In this section, we provide a detail overview of our traffic analytic architecture (see Figure 2, left). Our architecture consists of the following conceptual layers: the data layer, the processing layer and the application layer.

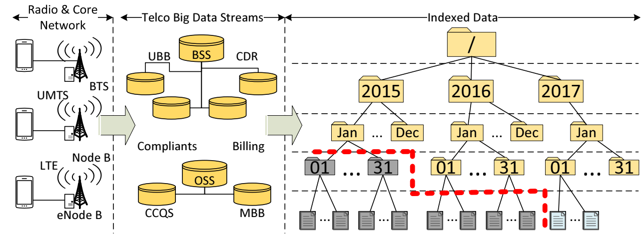

Figure 3: The life-cycle of telco big data. The telco network generates telco big data streams that are stored, indexed and analyzed by our system.

3.1 Data Layer

This layer is responsible to aggregate the incoming and stored data in an online and offline mode and provide the access methods to various input sources (e.g., telco data, map data, even social data). The data layer consists of a replicated big data file system in order to provide availability and performance. We have implemented a prototype with *Hadoop Distributed File System (HDFS)* [1] along with Apache Hive [2] and an RDBMS for catalog management. The architecture provides the required fault tolerance, scalability and efficiency. The index is created utilizing the folder-based structure of HIVE and HDFS, as shown in Figure 3. Thus, it can be scaled horizontally without any memory constrains. The data layer consists of the following two components (see Figure 2, left):

Storage: is responsible for minimizing the required storage space with minimal overhead on the query response time. The intuition is to use compression techniques that yield high compression ratios but at the same time guarantee small decompression times. We particularly use GZIP compression that offers high compression/decompression speeds, with a high compression ratio and has maximum compatibility with I/O stream libraries in the big data ecosystem we use [4].

Indexing: maintains and uses a multi-resolution spatiotemporal index that consists of the *Incremence* module and the *Highlights* module. The Incremence module receives the newly arriving snapshot and incorporates it into the index by incrementing it on the right-most path. The Highlights module combines data from the stored snapshots to create efficient representations of interesting events, called "highlights".

3.2 Processing Layer

This layer is providing the mechanisms for minimizing the query response time for data analytic queries. It is implemented on top of Apache Spark[3] for *online* (i.e., data-intensive distributed in-memory) data processing ensuring fast query execution time. The processing layer consists of the following four components:

Map Matching: is responsible to generate traffic points from the incoming sources and match them to the road network. In particular, map matching is expected to be based on shortest path [19].

[1]Apache HDFS: http://hadoop.apache.org/
[2]Apache Hive: http://hive.apache.org/
[3]Apache Spark: http://spark.apache.org/

Given that trajectories can be accurate within a 50 meter error bound [25], our first aim is accurately map these car trajectories to road segments using shortest path queries followed by drivers or by exploiting techniques off-the-shelf [27]. Secondly, given that a variety of trajectories emerge in a telco network (pedestrians, cars, trucks, etc.), our next aim is to do an activity recognition over TBD to adequately separate these users into classes. All of these aspects will require technical solutions, not addressed previously.

Privacy: is imperative to capitalizing on Traffic-TBD. Location data has a particularly high privacy requirement as it is easy to infer user activity from trajectories. This component aims to apply k-anonymity during the map-matching process so that no user-specific activity can be inferred within some statistical bounds. While k-anonymity protects against identity disclosure, it does not protect in general against attribute disclosure: if the values for one (or several) confidential attribute(s) are identical within a group of records sharing the quasi-identifier attribute values, disclosure happens. For this we aim to exploit more secure forms of k-anonymity, such as t-Closeness [14] or l-diversity [17], which imply more distortion to the original data than just achieving k-anonymity. The privacy module is related to the *Interactive Traffic Analytics* and the *API* (Open Data) of the Application Layer, presented in the next section. It either needs to be incorporated before any of these components or is incorporated separately within these two components. As stated in [8], its placement needs to take into account the trade-off between complexity of implementation and accuracy of the data produced.

Visualization: is providing the necessary data structures to efficiently pack and ship data to the analytical user interface. The module is responsible for continuously monitoring and the ability to "playback" over a selected past period in order to assist discovering interesting events. The rendering process of the heat map takes place on the client-side based on the partitioned data points, which allows dynamic interaction based on user input [4]. In our approach we also support heat map data that is saved as image tiles (e.g., PNG files) as opposed to actual data points, e.g., see Figure 2 (right), in order to cope with visualization scalability.

Mining: is responsible for finding interesting incidents through the historical and incoming data using the underlying distributed infrastructure. The module enables high performance event processing and real-time analytics on *Key Performance Indicators (KPIs)* to increase quality of service and enhance user experience. Examples of such queries include the identification of call drops due or slow network connectivity due to churning network traffic.

3.3 Application Layer

The application layer provides the interface with the end users. Our high level requirements, as these have been derived through various 1E-2.5E systems of similar scope, include the following queries:

A1. Hot Spot Analysis: Detect and visualize the most congested areas around a location.

A2. Navigation Monitoring: Calculate and visualize the best route based on the travel time and traffic queue length.

A3. Travel time: Calculate aggregated time loss due to congestion.

A4. Travel speed: Determine average speed based on the traffic flow.

A5. Accessibility analysis: Compute the time that is needed to reach a specific location at a specific time slot considering the road traffic.

A6. Incident detection: Detects any incidents due to traffic (e.g, car accidents, broken down vehicle).

A7. Before-after analysis: Measure the traffic flow after a change in the road network in order to be evaluated from the traffic administration.

A8. Travel behaviors: Identify travel behaviors throughout a region (e.g., napoleon diagrams or traffic flow diagrams at various time/space granularities).

These analytical tasks are expected to reach the users and developers through the following components: the *Interactive Traffic Analytics* component, the *Alert Service* component and the *API* component, as these are described below:

Interactive Traffic Analytics: receives a data analytic query and uses the index to combine the needed highlights to answer the query. The microanalytics provided by the *Interactive Traffic Analytics* component could allow some municipality to reduce traffic by more carefully planning the formation of its traffic network. It could even allow a transportation authority to simulate a blocked road (e.g., *"how would traffic evolve after events in certain areas?"*). Furthermore, road traffic administration will be able to act immediately without delays, utilizing the real-time analytics or the prediction of road congestion. This component is realized as an interactive map shown in Figure 2 (right), which is a custom web interface that has been implemented in Scala, HTML5/CSS3 along with extensive AngularJS and the Google Maps JS API and runs on the Play Framework v2.

Alert Service: is a background service that can provide real-time and predicted traffic alerts combining public APIs with the TBD (e.g., *"there will be traffic at road x"* or *"there is traffic at road x"*). For example, consider a bus that can be alerted for traffic on a specific road. The bus will be able take a different path dynamically to circumvent the traffic. The alert service can be also useful for the traffic administration (e.g., police), by informing them about traffic. This component is continuously interacting with the processing layer, especially with *Mining* and *Map Matching component* in order to have updated information. A preliminary version of this component has been implemented in Scala and it is running on the Play Framework v2 with the OneSignal push notification service.

API: makes the generated traffic insights available for public use. In particular, startups can use the open data to create innovative applications that can improve and support the navigation considering the road traffic. Furthermore, companies (e.g, insurance companies) can utilize the API to create marketing campaigns. The *API* component depends on the privacy component in the processing layer that guarantees privacy with some inherent accuracy trade-off. A preliminary version of the *API* component has also been implemented in Scala and runs over the Play Framework v2 providing a RESTful Web 2.0 API responding in JSON format.

4 CONCLUSIONS AND FUTURE WORK

In this position paper we outline the requirements and high-level design and building blocks of the Traffic-TBD architecture, which aims to become an innovative road traffic analytic and prediction system with the following desiderata: i) provide micro-level traffic modeling and prediction; ii) retain the location privacy boundaries of users inside their mobile network operator; and iii) be available with minimal costs and using existing infrastructure. In the future we are interested to address the inherent technical challenges pertinent to the following aspects: map-matching of trajectories with uncertain bounding envelopes, to provide privacy-preserving trajectory processing techniques, to exploit 2E/2.5E open traffic mapping data to learn traffic behaviors but also to experimentally implement our proposition in a real architecture and compare this architecture quantitatively against 1E and 2E systems.

ACKNOWLEDGMENTS

This work was supported in part by the University of Cyprus, an industrial sponsorship by MTN Cyprus and EU COST Action IC1304. The third author's research is supported by the Alexander von Humboldt-Foundation, Germany. The last author's research is supported by NSF grants IIS-0952977, IIS-1218168, IIS-1525953, CNS-1512877.

REFERENCES

[1] Javed Aslam, Sejoon Lim, Xinghao Pan, and Daniela Rus. 2012. City-scale Traffic Estimation from a Roving Sensor Network. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems (SenSys'12)*. ACM, New York, NY, USA, 141–154. https://doi.org/10.1145/2426656.2426671

[2] Eric Bouillet, Bei Chen, Chris Cooper, Dominik Dahlem, and Olivier Verscheure. 2013. Fusing Traffic Sensor Data for Real-time Road Conditions. In *Proceedings of First International Workshop on Sensing and Big Data Mining (SENSEMINE'13)*. ACM, New York, NY, USA, Article 8, 6 pages. https://doi.org/10.1145/2536714.2536722

[3] TomTom Intl. BV. 2017. TomTom Traffic. (2017). https://www.tomtom.com/

[4] Constantinos Costa, Georgios Chatzimilioudis, Demetrios Zeinalipour-Yazti, and Mohamed F. Mokbel. 2017. Efficient Exploration of Telco Big Data with Compression and Decaying. In *Proceedings of the 33rd International Conference on Data Engineering (ICDE'17)*. IEEE, San Diego, CA, USA, 1332–1343. https://doi.org/10.1109/ICDE.2017.175

[5] Preeti Goel, Lars Kulik, and Ramamohanarao Kotagiri. 2012. Privacy Aware Trajectory Determination in Road Traffic Networks. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems (SIGSPATIAL'12)*. ACM, New York, NY, USA, 406–409. https://doi.org/10.1145/2424321.2424376

[6] Qirong Ho, Wenqing Lin, Eran Shaham, Shonali Krishnaswamy, The Anh Dang, Jingxuan Wang, Isabel Choo Zhongyan, and Amy She-Nash. 2016. A Distributed Graph Algorithm for Discovering Unique Behavioral Groups from Large-Scale Telco Data. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM'16)*. ACM, New York, NY, USA, 1353–1362. https://doi.org/10.1145/2983323.2983354

[7] Shaohan Hu, Lu Su, Hengchang Liu, Hongyan Wang, and Tarek Abdelzaher. 2013. Poster Abstract: SmartRoad: A Crowd-sourced Traffic Regulator Detection and Identification System. In *Proceedings of the 12th International Conference on Information Processing in Sensor Networks (IPSN'13)*. ACM, New York, NY, USA, 331–332. https://doi.org/10.1145/2461381.2461433

[8] Xueyang Hu, Mingxuan Yuan, Jianguo Yao, Yu Deng, Lei Chen, Qiang Yang, Haibing Guan, and Jia Zeng. 2015. Differential Privacy in Telco Big Data Platform. *Proc. VLDB Endow.* 8, 12 (Aug. 2015), 1692–1703. https://doi.org/10.14778/2824032.2824067

[9] Yiqing Huang, Fangzhou Zhu, Mingxuan Yuan, Ke Deng, Yanhua Li, Bing Ni, Wenyuan Dai, Qiang Yang, and Jia Zeng. 2015. Telco Churn Prediction with Big Data. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (SIGMOD'15)*. ACM, New York, NY, USA, 607–618. https://doi.org/10.1145/2723372.2742794

[10] INRIX Inc. 2017. Los Angeles Tops INRIX Global Congestion Ranking. (2017). https://goo.gl/c2K3E1 Accessed 7-Aug-2017.

[11] Anand Padmanabha Iyer, Li Erran Li, and Ion Stoica. 2015. CellIQ: Real-time Cellular Network Analytics at Scale. In *Proceedings of the 12th USENIX Conference on Networked Systems Design and Implementation (NSDI'15)*. USENIX Association, Berkeley, CA, USA, 309–322. http://dl.acm.org/citation.cfm?id=2789770.2789792

[12] Ayush Jain, Saswat Raj, Harshit, Rajiv Misra, and B. M. Baveja. 2015. Road Congestion Sensing via Crowdsourcing and MapReduce. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks (IPSN'15)*. ACM, New York, NY, USA, 356–357. https://doi.org/10.1145/2737095.2742931

[13] Andreas Janecek, Danilo Valerio, Karin Anna Hummel, Fabio Ricciato, and Helmut Hlavacs. 2015. The Cellular Network as a Sensor: From Mobile Phone Data to Real-Time Road Traffic Monitoring. *Proceedings of the IEEE Transactions on Intelligent Transportation Systems* 16, 5 (Oct 2015), 2551–2572. https://doi.org/10.1109/TITS.2015.2413215

[14] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramaniann. 2007. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *Proceedings of the 23rd International Conference on Data Engineering (ICDE'07)*. IEEE, Istanbul, Turkey, 106–115. https://doi.org/10.1109/ICDE.2007.367856

[15] Kuien Liu, Ke Deng, Zhiming Ding, Mingshu Li, and Xiaofang Zhou. 2009. MOIR/MT: Monitoring Large-scale Road Network Traffic in Real-time. *Proc. VLDB Endow.* 2, 2 (Aug. 2009), 1538–1541. https://doi.org/10.14778/1687553.1687585

[16] Chen Luo, Jia Zeng, Mingxuan Yuan, Wenyuan Dai, and Qiang Yang. 2016. Telco User Activity Level Prediction with Massive Mobile Broadband Data. *ACM Trans. Intell. Syst. Technol.* 7, 4, Article 63 (May 2016), 30 pages. https://doi.org/10.1145/2856057

[17] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. 2007. L-diversity: Privacy Beyond K-anonymity. *ACM Trans. Knowl. Discov. Data* 1, 1, Article 3 (March 2007), 52 pages. https://doi.org/10.1145/1217299.1217302

[18] Cuong Pham and Nguyen Thi Thanh Thuy. 2016. Real-Time Traffic Activity Detection Using Mobile Devices. In *Proceedings of the 10th International Conference on Ubiquitous Information Management and Communication (IMCOM'16)*. ACM, New York, NY, USA, Article 64, 7 pages. https://doi.org/10.1145/2857546.2857611

[19] Mudhakar Srivatsa, Raghu Ganti, Jingjing Wang, and Vinay Kolar. 2013. Map Matching: Facts and Myths. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL'13)*. ACM, New York, NY, USA, 484–487. https://doi.org/10.1145/2525314.2525466

[20] Priyam Tejaswin, Rohan Kumar, and Siddharth Gupta. 2015. Tweeting Traffic: Analyzing Twitter for Generating Real-time City Traffic Insights and Predictions. In *Proceedings of the 2Nd IKDD Conference on Data Sciences (CODS-IKDD'15)*. ACM, New York, NY, USA, Article 9, 4 pages. https://doi.org/10.1145/2778865.2778874

[21] Arvind Thiagarajan, Lenin Ravindranath, Katrina LaCurts, Samuel Madden, Hari Balakrishnan, Sivan Toledo, and Jakob Eriksson. 2009. VTrack: Accurate, Energy-aware Road Traffic Delay Estimation Using Mobile Phones. In *Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems (SenSys'09)*. ACM, New York, NY, USA, 85–98. https://doi.org/10.1145/1644038.1644048

[22] Jie Xu, Dingxiong Deng, Ugur Demiryurek, Cyrus Shahabi, and Mihaela van der Schaar. 2015. Mining the Situation: Spatiotemporal Traffic Prediction With Big Data. *Proceedings of the IEEE Journal of Selected Topics in Signal Processing* 9, 4 (June 2015), 702–715. https://doi.org/10.1109/JSTSP.2015.2389196

[23] Shiming Zhang, Yin Yang, Wei Fan, Liang Lan, and Mingxuan Yuan. 2014. OceanRT: Real-time Analytics over Large Temporal Data. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data (SIGMOD'14)*. ACM, New York, NY, USA, 1099–1102. https://doi.org/10.1145/2588555.2594513

[24] Zimu Zheng, Dan Wang, Jian Pei, Yi Yuan, Cheng Fan, and Fu Xiao. 2016. Urban Traffic Prediction Through the Second Use of Inexpensive Big Data from Buildings. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM'16)*. ACM, New York, NY, USA, 1363–1372. https://doi.org/10.1145/2983323.2983357

[25] Pengfei Zhou, Shiqi Jiang, and Mo Li. 2015. Urban Traffic Monitoring with the Help of Bus Riders. In *Proceedings of the 35th International Conference on Distributed Computing Systems (ICDCS'15)*. IEEE, Columbus, OH, USA, 21–30. https://doi.org/10.1109/ICDCS.2015.11

[26] Tian Zhou, Lixin Gao, and Daiheng Ni. 2014. Road Traffic Prediction by Incorporating Online Information. In *Proceedings of the 23rd International Conference on World Wide Web (WWW'14 Companion)*. ACM, New York, NY, USA, 1235–1240. https://doi.org/10.1145/2567948.2580072

[27] Fangzhou Zhu, Chen Luo, Mingxuan Yuan, Yijian Zhu, Zhengqing Zhang, Tao Gu, Ke Deng, Weixiong Rao, and Jia Zeng. 2016. City-Scale Localization with Telco Big Data. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM'16)*. ACM, New York, NY, USA, 439–448. https://doi.org/10.1145/2983323.2983345