2016 17th IEEE International Conference on Mobile Data Management

# FoGBAT: Combining Bluetooth and GPS Data for Better Traffic Analytics

Johannes L. Borresen Christian S. Jensen Kristian Torp

Department of Computer Science
Aalborg University, Denmark
{jlb,csj,torp}@cs.aau.dk

***Abstract*—Congestion is a major problem in many cities. In order to monitor and manage traffic, a number of different sensor types are used to collect traffic data. This includes GPS devices in the vehicles themselves as well as fixed Bluetooth sensors along the roads. Each sensor type has advantages and disadvantages. Where GPS has a wide coverage of the road network, Bluetooth sensors gather data from a much higher number of vehicles. In this paper we present FoGBAT, a system that combines GPS data with Bluetooth data. The goal of the system is to retain the advantages of both. We show how the data types are aligned to ensure that data from each sensor type is related to the exact same part of the road network and cover the same time period. Using very large real-world data sets, we use the system to compare travel speeds based on each data type, and how the use of both data types simultaneously can improve the accuracy of computed travel speeds and congestion levels.**

## I. Introduction

Road traffic is increasing steadily, with transportation infrastructures being unable to keep up with the increase, which results in increased congestion and reduced mobility [1]. This situation may be addressed by more intelligent use of existing infrastructure and by making purposeful improvements to infrastructures. These in turn should be based on accurate, detailed, and up-to-date traffic data.

With the growing deployment of online GPS receivers, e.g., in smartphones and in in-vehicle navigation systems, GPS data is a valuable source of vehicular traffic data: When timestamped GPS positions of vehicles carrying GPS receivers are captured sufficiently frequently, this data captures the trajectories of the vehicles, which offers detailed insight into the movements of the vehicles; and even less frequent GPS data from a vehicle offers valuable information on the vehicle's movement [2].

However, the use of GPS data as a means of capturing traffic also faces challenges: It is possible to capture the traffic on a particular road segment at a particular time only if data is available from a sufficient number of vehicles that traversed the segment at the given time. Data from a single vehicle may not be representative.

With the data available to us, we are able to cover all major roads, as opposed to residential roads, in the road network of Denmark with some GPS data. However, for many road segments and times of day, we have data from relatively few vehicles. Because it is difficult to get data from a large fraction of all vehicles traveling in an infrastructure, even with continued data collection, this situation is likely to persist. In sum, it is possible to cover all major roads with GPS data, but only little data is often available for a segment and a particular time.

Road authorities have started to deploy roadside equipment, notably Bluetooth boxes, that detect Bluetooth devices when these pass by the boxes. Thus, when a smartphone in a vehicle is connected to the vehicle via Bluetooth, the vehicle is seen by this equipment. By deploying pairs of boxes, it is possible to capture time-varying travel times between these boxes. This data has different characteristics than GPS data. On the one hand, it captures only the traffic on a few road segments due to the deployment costs. On the other hand, it is capable of capturing a substantial fraction of all traffic on the segments where it is deployed.

There are many potential benefits to integrating GPS and Bluetooth data: (i) GPS data can extend the limited coverage achieved by Bluetooth data. (ii) Bluetooth data can capture traffic in more detail where it is available and thus is preferable over GPS data when it is available. (iii) Bluetooth data can be regarded as a ground truth and can be used to assess the inaccuracies and uncertainties caused by the limited amount of GPS data on some segments and at some times.

With benefits such as these in mind, we present a system that enables traffic analytics on combined GPS and Bluetooth data in order to exploit the complimentary strengths of the two kinds of data. FoGBAT allows users to explore how the two kinds of data complement each other along a number of temporal factors. Previously, Patire et al. [3] have studied the benefit of combining GPS and Bluetooth data in order to predict travel times. FoGBAT, instead, focuses on the exploration of average travel speeds through user-defined temporal filters. To the authors' knowledge, this is the first such application of Bluetooth and GPS data.

The remainder of the paper first presents the two data sources used. This is followed by an overview of the FoGBAT system, describing its components and the key challenges addressed by each component. Finally, an overview is given of the features that will be demonstrated to conference participants.

2375-0324/16 $31.00 © 2016 IEEE
DOI 10.1109/MDM.2016.55

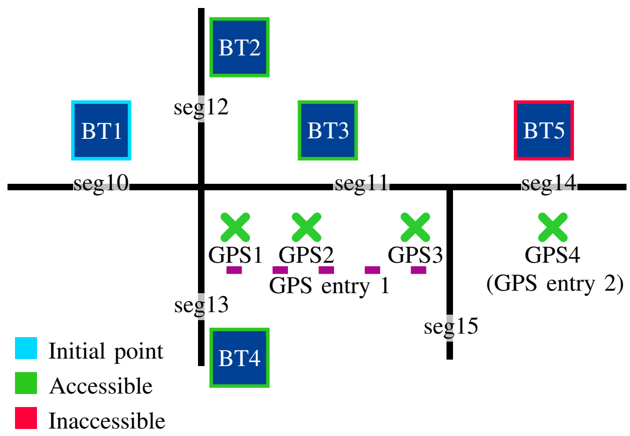


Fig. 1. Example of Bluetooth and GPS Data Collection

TABLE I
EXAMPLE OF GPS RECORDS

| recordid | vehicleid | time | lat/lng | speed |
|---|---|---|---|---|
| GPS1 | 42 | 7:45:10 | - | 15 |
| GPS2 | 42 | 7:45:15 | - | 25 |
| GPS3 | 42 | 7:45:20 | - | 40 |
| GPS4 | 42 | 7:45:25 | - | 50 |

TABLE II
EXAMPLE OF BLUETOOTH RECORDS

| from_time | to_time | from_box | to_box | count | speed |
|---|---|---|---|---|---|
| 7:45 | 7:50 | BT1 | BT2 | 30 | 30 |
| 7:50 | 7:55 | BT1 | BT2 | 50 | 35 |
| 7:50 | 7:55 | BT1 | BT4 | 90 | 45 |

## II. DATA SOURCES

A GPS receiver emits so-called NMEA sentences that are records with a range of fields, some of which are included in the GPS records that we use. Specifically, a GPS record contains a vehicle identifier, a time, latitude/longitude coordinates, and a speed. Such records are aggregated into combined records, one for each OpenStreetMap (OSM) road segment traversed. An aggregated record is a tuple $t = (seg, time, avg)$, containing the road segment, the earliest time of the single records, and their average speed. Figure 1 shows an example using the data shown in Table I. The four green crosses represent GPS records as a vehicle accelerates away from an intersection, from left to right. The three points *GPS1*, *GPS2*, and *GPS3* belong to road segment *seg11*, while *GPS4* belongs to *seg14*. Combined, this yields two aggregated records, $(seg_{11}, 7{:}45{:}10, 26.667)$ and $(seg_{14}, 7{:}45{:}25, 50)$.

Next, Bluetooth data consists of average traversal speeds for segments between pairs of Bluetooth boxes for 5-minute intervals. Figure 1 contains an example. Here, box *BT1* is paired with each of the boxes *BT2*, *BT3*, and *BT4*. Thus, a vehicle that is seen at *BT1* and then at *BT2* contributes to an aggregate travel time for the segment between *BT1* and *BT2* during a 5-minute period. Table II contains examples of records as they might appear in pairings starting from *BT1*. Each row describes a 5-minute period for a pair of Bluetooth boxes, the total number of vehicles registered for the period along with their average traversal speed.

Note that the segments captured by Bluetooth boxes do not align with the OSM road segments. For example, the Bluetooth pair *(BT1, BT3)* covers the right side of *seg10* and the left side of *seg11*, while *(BT3, BT5)* covers the right side of *seg11* and the left side of *seg14*.

## III. SYSTEM OVERVIEW

Figure 2 shows an overview of the FoGBAT system. Light blue rectangles are modules, while a light red rectangles with rounded corners are data sources. The core modules are those for *stretch identification*, *data integration*, and *visualization*. FoGBAT uses a versatile software stack that provides a number of functionalities [2], [4]–[7].

A pair of Bluetooth boxes containing an identifier of the pair as well as the coordinates for the starting and ending boxes is given as a tuple $(pid, from\_latlon, to\_latlon)$. These coordinates, while accurate, do not map directly to the segments of a road for which the pair captures traffic. The *stretch identification* module computes the segments between pairs of boxes, making it possible to compare the captured Bluetooth data with the GPS data for those segments. The module applies Dijkstra's algorithm as implemented in the pgRouting extension [8] for the PostgreSQL database system to map each pair to a road segment. Such a path between two Bluetooth boxes is called a *stretch*.

Although Bluetooth boxes have been mapped to road segments, Bluetooth records are not yet comparable to GPS data. Recall that a Bluetooth stretch does not align with the OSM road segments for GPS data. Given a Bluetooth stretch, the *data integration* module retrieves only the GPS records that are map-matched to the road segments that constitute the Bluetooth stretch. Further, the module aggregates data to the user-defined temporal granularity. If granularities more coarse than the 5-minute periods of Bluetooth data are requested, the module computes a weighted average from the default 5-minute Bluetooth periods. Referring to the examples in Table II, the module might produce a record for a 10-minute period by combining the records for 7:45 to 7:50 and 7:50 to 7:55 for the Bluetooth stretch starting at *BT1* and ending at *BT2*:

$$\left(7{:}45, 7{:}55, \mathrm{BT1}, \mathrm{BT2}, 30 + 50, \frac{30 \cdot 30 + 50 \cdot 35}{30 + 50}\right)$$

The result is that GPS and Bluetooth data can now be compared directly, as they cover the exact same part of the road network and the same temporal granularity of time.

Finally, the *visualization* module produces different views of the integrated data. The start page of the web-based application shows all Bluetooth strethes as overlays on an OSM map

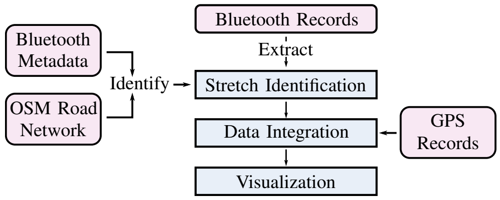


Fig. 2. System Overview

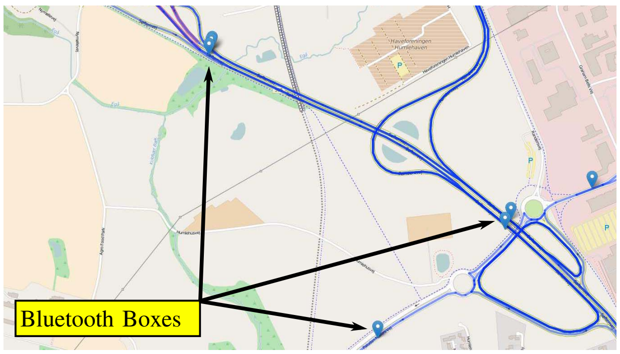


Fig. 3. Screenshot of Stretches in FoGBAT

(Figure 3). If a particular stretch is queried, optionally with a number of temporal filters, the module generates a number of different views on the output from the data integration module. Examples of these views are shown in Figure 4 and explained in greater detail in the next section.

FoGBAT additionally serves as a platform to further explore the relationship between GPS and Bluetooth data. A few possibilites are described briefly here.

*Granularity of Time:* It is of general interest to examine changes at the granularity of minutes in traffic, to better respond to anomalies as they arise. For example, it would be preferable to respond to accidents as soon as possible. For GPS data, such a fine granularity is sensitive to outliers, while Bluetooth data is restricted to a 5-minute period. FoGBAT enables the exploration of this relationship across many granularities, describing the properties of a finer GPS granularity when compared to Bluetooth data.

*Confidence of non-stretch GPS aggregates:* As mentioned earlier, the Bluetooth data originates from a much larger set of vehicles than the GPS data. Thus, the Bluetooth data can serve as ground truth data for the segment's traffic, and correlation of the Bluetooth data with the GPS data for a section can offer insight into the accuracy and uncertainty of the GPS data. The knowledge can then be extrapolated to estimate the accuracy and uncertainty of similar segments where only GPS data is available.

*Efficient expansion of Bluetooth infrastructure:* Given the costs of new infrastructure, it is pertinent to consider where new Bluetooth boxes will offer the most valuable new insight into the traffic flow. For example, if a road segment is heavily used, but is only passed by few vehicles with GPS devices, it

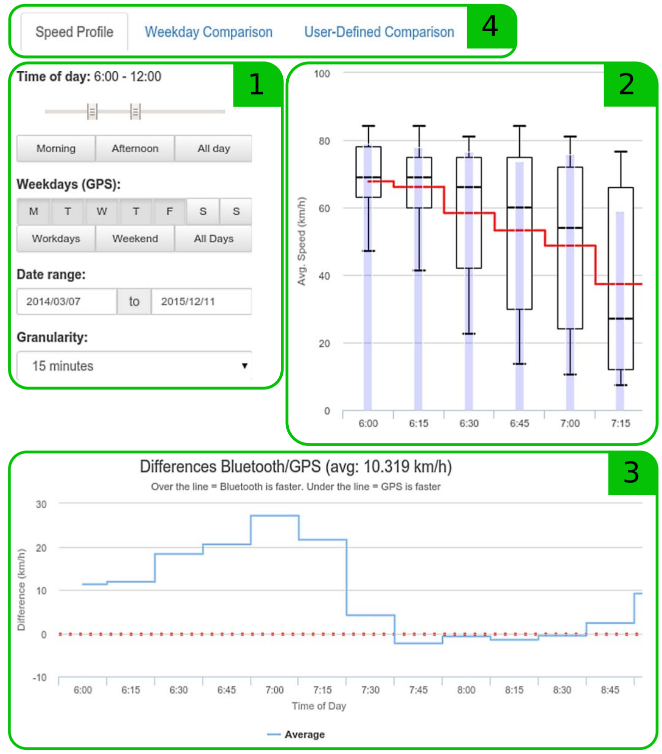


Fig. 4. Screenshot of FoGBAT's Stretch Details

would benefit from a pair of Bluetooth boxes. FoGBAT can assist traffic managers in determining which stretches of road will gain the most from a new pair of Bluetooth boxes.

## IV. Demonstration Outline

This section outlines the use cases that the conference participants will be able to explore. FoGBAT's main interface is a map that shows which stretches of road are covered by Bluetooth data (Figure 3). Upon selecting a stretch, a frame opens that enables a number of analyses.

By default, the tab with the *Speed Profile* is shown (Figure 4), which compares the Bluetooth and GPS data selected by the temporal filters in the frame's left-hand-side panel (Label 1 in Figure 4); similar filters are available for the analyses in the other tabs. The top chart (Label 2 in Figure 4) shows average speeds (km/h) for Bluetooth (blue bar) and GPS (red plot line) data, with an additional boxplot for the GPS data. Underneath, a second chart displays the differences between the GPS and Bluetooth averages (Label 3 in Figure 4). A red, dotted line denotes a neutral center. If the blue line follows the center, the GPS and Bluetooth averages are equal.

At the top of the display (Label 4 in Figure 4), two additional comparisons are available.

The *Weekday Comparison* (Figure 5) breaks Bluetooth data into single weekdays, and compares these to a GPS average. By default, an aggregate of Monday through Friday is presented for the GPS data. Again, a number of filters on the left hand side are available to tailor the query.

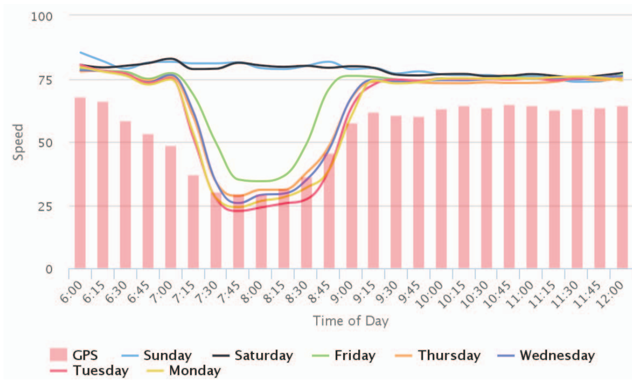


Fig. 5. Weekday Comparison

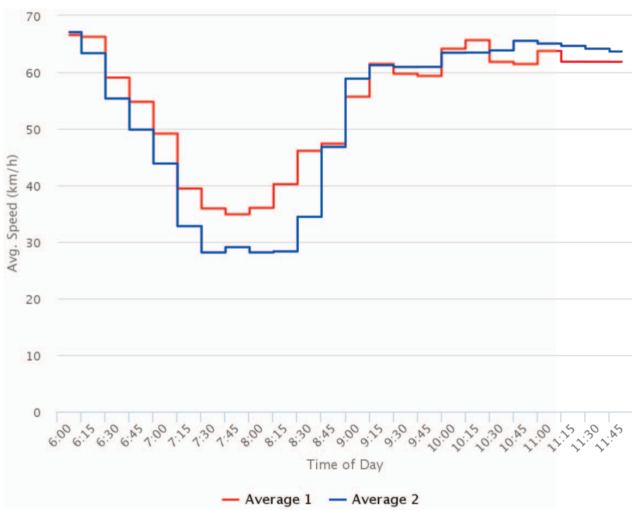


Fig. 6. User-Defined Comparison

Finally, the *User-Defined Comparison* (Figure 6) allows for user-defined comparisons between the GPS and Bluetooth data sets. This includes comparisons between different periods for the same data source. In in Figure 6, GPS data for the winter months of 2014/2015 (in blue) are compared to GPS data for the summer months of 2015 (in red). Notice how vehicles generally travel slower during winter.

Conference participants will be able to explore the currently integrated data sets in the Danish cities of Aalborg and Aarhus.

## V. Related Work

A number of papers have addressed the challenge of combining traffic data from multiple types of sources [9]–[11]. However, these papers use either limited data sets or synthetic data. Patire et al. [3] propose a framework for combining multiple sensor types with a focus on achieving improved travel-time accuracy. They study how GPS data is needed to achieve accurate travel-time estimates.

Janssen et al [12] combine induction-loop and Bluetooth data to get information about bicycle traffic in cities. Wang et al. [13] detect congestion by combining information from Twitter micro-blog posts with historical travel-time data. Zheng et al. [14] also use a combination of social-media and traffic data with the goal of determining noise level in a city.

## VI. Summary

We have shown how FoGBAT integrates the visualization of vehicular Bluetooth and GPS data, enabling analyses of their relationship along road segments. Apart from expanding the geographical coverage of the underlying data, it is of particular interest to integrate new data sources into the system to enable studies of their interplay. This can include new types of data, such as from automatic license plate readers, or supplementary GPS data. Overall, this may enable improved modeling of traffic and predictions of traffic flow.

## Acknowledgments

This research was supported in part by the DiCyPS center, funded by Innovation Fund Denmark, and a grant from the Obel Family Foundation.

## References

[1] J. Dulac, "Global land transport infrastructure requirements," *International Energy Agency, Paris*, 2013.
[2] O. Andersen, B. Krogh, and K. Torp, "An open-source based ITS platform," in *Proc. 14th IEEE International Conference on Mobile Data Management*, vol. 2, 2013, pp. 27–32.
[3] A. D. Patire, M. Wright, B. Prodhomme, and A. M. Bayen, "How much GPS data do we need?" *Transportation Research Part C: Emerging Technologies*, vol. 58 part B, pp. 325–342, 2015.
[4] O. Andersen, C. S. Jensen, K. Torp, and B. Yang, "EcoTour: reducing the environmental footprint of vehicles using eco-routes," in *Proc. 14th IEEE International Conference on Mobile Data Management*, vol. 1, 2013, pp. 338–340.
[5] C. Guo, B. Yang, O. Andersen, C. S. Jensen, and K. Torp, "EcoSky: reducing vehicular environmental impact through eco-routing," in *Proc. 31st IEEE International Conference on Data Engineering*, 2015, pp. 1412–1415.
[6] ——, "Ecomark 2.0: empowering eco-routing with vehicular environmental models and actual vehicle fuel consumption data," *GeoInformatica*, vol. 19, no. 3, pp. 567–599, 2015.
[7] B. Krogh, O. Andersen, E. Lewis-Kelham, and K. Torp, "Efficient one-click browsing of large trajectory sets," in *Proc. 22nd ACM International Conference on Advances in Geographic Information Systems*, 2014, pp. 569–572.
[8] pgRouting Project. [Online]. Available: http://pgrouting.org/
[9] W. Shen and L. Wynter, "Real-time road traffic fusion and prediction with GPS and fixed-sensor data," in *Proc. 15th IEEE International Conference on Information Fusion*, 2012, pp. 1468–1475.
[10] Y. Feng, J. Hourdos, and G. A. Davis, "Probe vehicle based real-time traffic monitoring on urban roadways," *Transportation Research Part C: Emerging Technologies*, vol. 40, pp. 160–178, 2014.
[11] A. Nantes, D. Ngoduy, A. Bhaskar, M. Miska, and E. Chung, "Real-time traffic state estimation in urban corridors from heterogeneous data," *Transportation Research Part C: Emerging Technologies*, 2015 (in press).
[12] S. Jansen, D. Hoting, J. Runge, T. Brinkhoff, D. Nicklas, and J. Sauer, "9 million bicycles? Extending induction loops with Bluetooth sensing," in *Proc. 15th IEEE International Conference on Mobile Data Management*, vol. 1, 2014, pp. 113–116.
[13] S. Wang, L. He, L. Stenneth, P. S. Yu, and Z. Li, "Citywide traffic congestion estimation with social media," in *Proc. 23rd ACM International Conference on Advances in Geographic Information Systems*, 2015, p. 34.
[14] Y. Zheng, T. Liu, Y. Wang, Y. Zhu, Y. Liu, and E. Chang, "Diagnosing New York city's noises with ubiquitous data," in *Proc. 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014, pp. 715–725.