2010 13th International IEEE
Annual Conference on Intelligent Transportation Systems
Madeira Island, Portugal, September 19-22, 2010

Requirements and Potential of GPS-based Floating Car Data for Traffic Management: Stockholm Case Study

Mahmood Rahmani, Haris N. Koutsopoulos, Anand Ranganathan

Abstract—**The application of GPS probes in traffic management is growing rapidly as the required data collection infrastructure is increasingly in place in urban areas with significant number of mobile sensors moving around covering expansive areas of the road network. The paper presents the development of a laboratory designed to explore GPS and other emerging traffic and traffic-related data for traffic monitoring and control. It also presents results to illustrate the scope of traffic information that can be provided by GPS-based data, using the city of Stockholm as a case study. The preliminary analysis shows that network coverage, especially during peak weekday hours, is adequate. Further investigation is needed to validate the data, and increase its value through fusion with complementary data from other sources.**

I. INTRODUCTION

INTELLIGENT Transportation Systems (ITS) have brought many advances in the transportation field. An important development is the emergence and installation of sensor technologies for the collection of data on the state of the transport system. As a result of ITS, extensive data collection infrastructures have been developed and even implemented in many urban areas collecting traffic data with various technical characteristics, including type of data, accuracy, and network coverage. In addition, weather and environmental data are also collected. However, at this early stage of development, many of the data are collected by separate entities and therefore remain fragmented and not integrated.

GPS is an excellent example of this new generation of sensors. GPS is also an example of opportunistic sensors that have the potential to provide high quality traffic data for real-time traffic monitoring and management, as well as planning, policy, and services, at a relatively low cost. GPS data, or more generally floating car data (FCD), represent the location of vehicles collected by mobile sources using GPS devices installed in vehicles. In many cases the raw data is transmitted to a central facility for processing. In other cases, advanced GPS equipment with built-in digital road map carry out map-matching at the same time so that the system is capable of reporting the link name and trajectory at once, as is the case of navigation equipment, [1].

The potential of GPS-based traffic data collection methods is even greater in light of the deployment of GALILEO [2] that will offer higher accuracy than current alternatives. In fact, transport has been one of the most important application areas for GALILEO (spanning traffic, bus, rail, air, and maritime transport).

Wireless phone service providers can automatically collect geo-location data, which can then be used to extract flow and speed information, and provide another source of FCD. The concept is still at an experimental stage, but applications are emerging, e.g. measurement of traffic speeds and travel times [3], identification of spatial and temporal congestion characteristics [4], etc.

The Mobile Millennium project aimed at a proof of concept for using smart phones with GPS to collect traffic data and use them for real-time traffic estimation and prediction. The results from a large-scale freeway and arterial experiment are very encouraging and highlight the significance of FCD for traffic management [5].

FCD is not limited to car data only. Technologies such as Automated Vehicle Location (AVL) have been in use in the transit industry for a while. Modern AVL systems are based on GPS or differential GPS, often augmented by dead reckoning, for collection of data on vehicle location, speed and other information [6].

Because of the potential of FCD, auto manufacturers have also been involved in its further development to include vehicle status data, such as ABS signals, headlight status, speed, acceleration, distance from surrounding objects, status of crash sensors, etc [7]. Built-in detection algorithms filter out false data (e.g. stopping to pick up passengers). In another extended FCD, called, CityFCD, the travel time is measured on-board and a message is transmitted to a traffic monitoring center only if it differs from the default link time [8]. This reduces frequency of message transmission by a factor of 40.

In general, FCD has a lot of advantages since it is cost effective, almost continuous, with potentially extensive spatial coverage. As such, FCD has the potential to contribute to many applications: monitoring and control of the traffic system, including queue and incident detection, dynamic route guidance, real-time multimodal information for travelers, short-time forecasting [1].

On the other hand, FCD may occasionally have low accuracy (including blind spots in the network). Depending

Manuscript received April 19, 2010.

Mahmood Rahmani, PhD student, Department of Transport and Economics, The Royal Institute of Technology, KTH. Teknikringen 72 SE - 100 44 Stockholm, SWEDEN (phone: 0046-8-790-8301; fax: 0046-8-21-2899; e-mail: mahmoodr@ kth.se).

Haris N. Koutsopoulos, professor, Department of Transport and Economics, The Royal Institute of Technology, KTH. Teknikringen 72 SE - 100 44 Stockholm, SWEDEN (e-mail: hnk@abe.kth.se).

Anand Ranganathan, Research Staff Member, IBM TJ Watson Research Center, 19 Skyline Drive, Hawthorne, NY 10532, USA (email: arangana@us.ibm.com).

on the source, in some cases, the frequency of probes is low, and that makes preprocessing (e.g. map-matching) more difficult. Traceability associated with FCD causes public concerns about privacy.

Furthermore, GPS currently provides relatively raw data that need to be appropriately processed. Further research and development is needed to make the data directly usable in real world traffic applications, [1].

The objective of this paper is twofold:

a) present the development of an ITS laboratory with novel computational capabilities based on the stream processing computing paradigm. The laboratory is designed to support real-time processing of data from emerging traffic data sensor technologies. The targeted applications include traffic information generation and multimodal travel planning by individual and fleet managers, as well as a host of other applications related to monitoring and control of transport systems; and
b) investigate, through a case study, the potential of FCD in terms of temporal and spatial coverage of an urban area, and explore the main attributes of data typically provided by GPS sensors.

The paper is organized as follows. Section II describes the prototype ITS laboratory. Section III presents results related to one type of data the lab is dealing with (GPS probe data), and section IV concludes the paper.

II. ITS Laboratory For Real-time Traffic Data Processing

With the recent developments in ITS, a multitude of data sources is rapidly becoming available. The availability of data from the various emerging sources, including GPS, offers many opportunities for monitoring, management, and control of traffic operations. However, in many cases the data is not integrated and opportunities for improved applications and services are not taken advantage of. Only through integration of the various data sources a more complete and accurate picture of traffic and travel conditions in a region, both in real-time and historically, can be obtained.

To accomplish this objective an ITS Laboratory has been developed at KTH, in Stockholm. The objective is to both develop the functionalities required; and demonstrate the use and value of the traffic data currently available and expected to become available routinely in the future. The lab combines three important elements as Figure 1 illustrates: data infrastructure, computing infrastructure and information infrastructure.

a. Data Infrastructure

A number of sources collect traffic data and traffic related data in the Stockholm region. Many of these data is currently received at the ITS Lab in real-time, with plans to receive the remaining of the data in the near future.

GPS. The GPS data is mainly from taxis. Public transport and select commercial vehicles also report their locations.

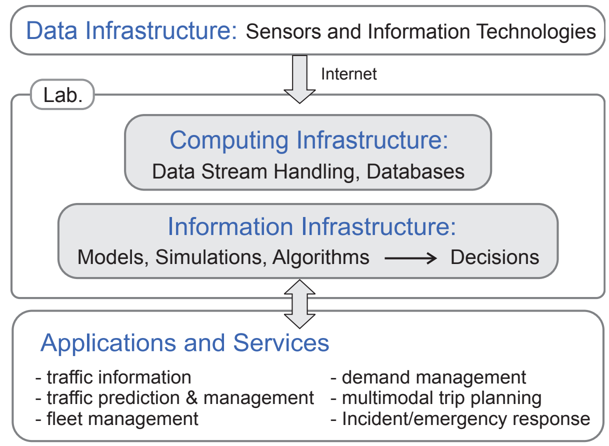

Fig. 1. ITS Laboratory elements and applications

The rate of reading and other statistics of the GPS traces will be discussed in section III.

Motorway Control (MCS). The Motorway Control System uses a number of microwave sensors placed in parts of the network to collect flow and speed data per lane, aggregated in 1-minute intervals.

Road data. Road data includes accident information, road works, etc. The data include timestamp, location, and estimated duration of the event.

Congestion pricing data. Stockholm has a time-of-day congestion pricing system. The system is based on plate recognition through video technology. The 18 detection stations, strategically located at crossing points provide data on flows, by time of day.

Travel time data. License plate recognition based on video processing is used not only by the congestion pricing system, but also by the city to collect travel time data between select intersections. The system is deployed to support development of real-time travel time information for major arterials in the Stockholm downtown area.

Public Transport. Buses are equipped with AVL that monitor the performance of the system and broadcast related information (including for example, how buses are running relative to the schedule).

Environmental data. A network of stationary sensors continuously collects data related to emissions at critical intersections in the system.

Road Weather Information System (RWIS). RWIS provides information about weather conditions as related to roadways such as temperature, rain/snowfall, black ice, wind, etc.

Other. Other traffic data, typical of legacy systems are also available, such as counts and speeds from loop detectors.

b. Computing Infrastructure

Studies have shown that developing and integrating the various components of an ITS infrastructure constitute a significant portion of the capital cost and complexity of such systems [9]. Furthermore, handling the massive data from a very large number of vehicles efficiently can be challenging for large-scale real-time traffic applications. Therefore, integration and processing of the traffic data from diverse sources requires a computing platform that is scalable and flexible to support the diverse needs of the various applications (ranging from real-time monitoring and control to archiving and long-term planning), and facilitates development and implementation.

The computing platform should be able to handle data streams from all relevant sources (motorways, urban streets, tunnels, transit, rail, weather, environmental sensors, etc) and associated databases. *Stream processing* is a computing paradigm that has the potential to satisfy these needs. The main requirements of real-time stream processing include [10]: keeping the data moving, having SQL capability on streams, handling stream imperfections, predictable outcome, high availability, stored and streamed data handling, distribution and scalability.

The IBM *InfoSphere Streams* [11] platform is used to provide the stream computing capabilities for the ITS Laboratory. It scales applications to a large number of compute nodes organized as a shared-nothing cluster of workstations or as a large supercomputer. A data-flow graph consists of a set of operators connected by streams, where each stream carries a series of Stream Data Objects. Each operator implements data stream analytics and resides in execution containers called Processing Elements, which are distributed over the compute nodes. The operators communicate with each other via their input and output ports, connected by streams.

InfoSphere Streams supports a declarative language, SPADE [12], to program stream-processing applications and to define the data-flow graph. SPADE supports interfaces with predefined operators and user-defined operators (in C++ or Java).

InfoSphere Streams includes a scheduler component that decides how to partition the processing of the data across a distributed set of physical nodes [13]. The scheduler uses the computational profiles of the operators, the loads on the nodes and the priority of the application in making its scheduling decisions.

InfoSphere Streams has been used to analyze the GPS data received in the ITS Lab [14]. The application performs different tasks such as cleaning the incoming GPS data, matching it to a map, generating statistics per link and per region, estimating travel times between different pairs of points, etc (Figure 2).

c. Information Infrastructure

The available data and the stream computing platform support the implementation of a number of operations on the data to extract information that is most useful for the intended application. Such analytics may include short-term traffic prediction, calculation of (stochastic) shortest paths, aggregation for archiving, calculation of congestion indices

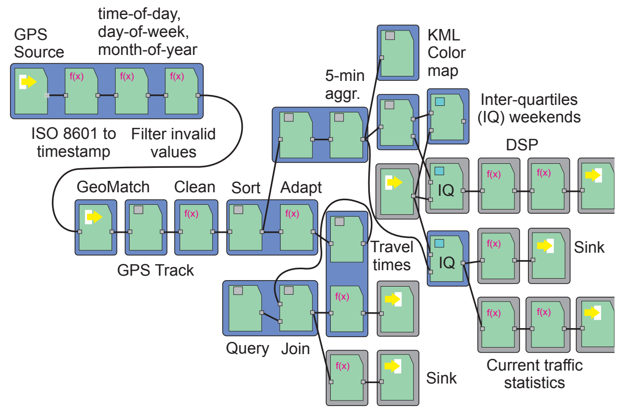

Fig. 2. A sample stream processing application used to analyze the GPS data.

using all available information, etc. Traffic simulation models, at various levels of resolution and granularity, can also be combined to complement the needed functionality.

Based on the structure described above and the data streams obtained in real-time, the ITS Lab facilitates a number of applications spanning operations, evaluation, and monitoring and control of transport systems. Examples include:

- *Information generation and trip planning.* The information can support a number of applications at higher levels of precision than is currently available:
  - door-to-door multimodal trip planning
  - traffic information for fleet management

  Information about the reliability of the various alternative modes can also be included.
- *Testing and evaluation.* Wide range of new technologies, systems, and concepts for traveler information services, advanced transportation and fleet management, intermodal services/facilities from an operational point of view, and demand management concepts (for example, new congestion pricing strategies), etc.
- *Reporting and performance monitoring.* The ITS Lab, organized in the way discussed here, can also help the monitoring of the system performance over time by generating high level information appropriate for policy makers. For example, the data collected can be used to estimate aggregate congestion performance measures and help publish related reports for the region.
- *Travel time variability and other measures of system performance.* Travel times and travel time variability is an important measure of quality related to the mobility of the transport system in urban areas.
- *Archived data for planning and evaluation.* The lab can serve as the repository of traffic and transportation related data, properly archived. Archived information is also useful to identify trends, changes in patterns, and

point to problem areas, e.g. bottlenecks, etc.

III. Stockholm Taxi GPS Data Analysis

There is a lot of research activity in the use of GPS data for traffic applications. Research focuses on a number of areas with two important ones related to potential applications for traffic management; and processing of the raw location data provided by GPS and analysis of their accuracy.

The OPTIS project ([15]) tested FCD in a field trial in Gothenburg. 250 cars were involved in this study during a period of 6 months. The results of OPTIS show that high quality travel information can be produced from GPS data. The installation cost of the FCD solution was estimated to be half of that of a stationary detector system.

The Institute of Transport of the German Aerospace Center reported ([16]) that GPS-based FCD produced by several hundreds taxis could result in an almost complete coverage of all major roads in the urban areas of Berlin, Nuremburg, and Vienna (with 300, 500, and 600 taxis respectively). The probes were transmitted at intervals of between 15–120 seconds. Based on the analysis of the GPS data, the study concluded that there are differences between the cities in terms of speeds and daily variations. It was also reported that significant decrease in speeds is associated to events such as bad weather conditions or roadwork.

In a related study travel time data were collected in Nuremberg in order to evaluate the FCD collected by 500 participating taxis ([17]). The result showed that FCD is able to detect congested situations and provide reliable travel time information. The study also suggested that fusion with data from other sensors can improve the performance of the system.

A recent study in Japan ([18]) examined the feasibility of using taxis as probe vehicles. The results indicated that the frequency of probes was not high enough in order for the data to be useful. When the time between consecutive probes from the same vehicle is large, the accuracy of link travel time measurements suffers, but combining FCD with historical data can help overcome the problem.

In order for FCD to be useful, it needs to be matched with the corresponding digital road network. Map matching is a critical preprocessing step. The accuracy of both probes and digital maps, and frequency of probes are crucial to a map matching process. Map-matching of low frequency probes (i.e. one per 1-5 minutes), especially in dense networks, is difficult [19]. In [20] map-matching algorithms for transportation applications are examined in detail.

The FCD provided by taxis in Stockholm includes time-stamped location (latitude and longitude), and status (free or meter-on), and received anonymously. The feasibility of using this data, fused with the traffic data from other sources, for traffic monitoring, real-time traffic information, traffic management, and dynamic journey planners, is of great interest. As a first step towards this, an analysis of the data in terms of their spatial and temporal coverage of the network is reported. An initial investigation of the travel time data collected through the probes is also discussed.

a. Probe Frequency

The ITS lab receives GPS probes from 1500 taxis working in Stockholm. The total number of probes per month is about 10 million.

A time lag, defined as the difference between the receipt and measurement times, is considered as one of the quality measures of the data received by the probes. This measure shows *how old the data is* when it is received at the lab. A broker, who gathers the data from the taxi network and delivers it to the ITS lab, delivers the data in batches at a frequency of one batch per minute. That means the probes are being received with latency of a few seconds up to one minute. Figure 3-left depicts the cumulative distribution of the time lag for the period of 1st to 5th of March 2010. Probes more than 3 standard deviations away from the average are considered outliers. The data is on the average 39 seconds old by the time they arrive at the lab.

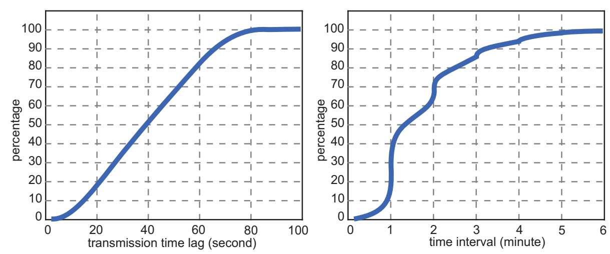

Fig. 3. Cumulative distribution for the transmission time lag (left) and the time interval between probes of each vehicle (right).

Another quality factor of GPS probes is *the frequency of transmission* from each individual vehicle. Figure 3-right shows the cumulative distribution for the time between consecutive probes from the same vehicle. The transmission frequency depends on the status of each taxi (occupied, free, etc.), and is about 110 seconds on average.

b. Temporal Coverage

The total number of probes reported in a period of time depends on the number of active (engaged with a client) taxis in that period and the frequency of reporting probes by taxis. Number of active taxis depends on the level of demand for taxi in the city, which varies by the time of a day. The two known traffic peaks, one in the morning and the other in the evening of working days, have the most demand for taxi in the city; hence one expects large number of probes reported by taxis during the peak hours.

Figure 4 depicts the number of probes aggregated in 15-minute intervals, between 6:00 and 20:00, from 18th January until 8th March 2010. It indicates that weekdays follow similar patterns in terms of variation of the number of probes by time of the day. The demand for taxis during weekends is generally less than working days, except for Saturday evening. The plot shows the two peak periods for weekdays, indicating that the morning peak is sharper, while the afternoon peak is wider but not as high. The data also show a

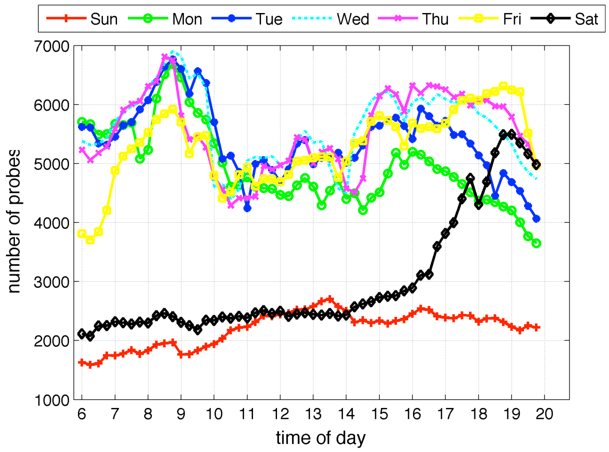

Fig. 4. Number of probes in 15 min intervals, 6:00-20:00, 18th January-8th March 2010.

similar pattern among the weekdays, especially during the morning peak. Afternoon peaks exhibit slightly different patterns.

Figure 5-left illustrates the average number of probes within a week, from data collected in the same period. It also illustrates the variability in the number of probes with Thursdays exhibiting the highest variability.

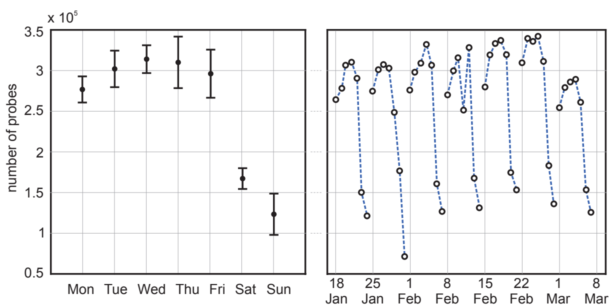

Fig. 5. Left: average ± 1 standard deviation of the number of probes per day of the week, 18th Jan - 8th Mar 2010. Right: number of probes per day.

Figure 5-right shows the number of probes per day for that period. The data show that usage, in general, repeats itself on a week-by-week basis. However, there are few weeks with slightly increased usage that coincides with a period when extreme weather conditions created a lot of problems and cancelations in the public transport system in the region (2nd half of February 2010).

c. *Spatial Coverage*

Stockholm, with 1500 probing taxis has relatively good penetration rate, compared to other studies referred in section III. As the analysis that follows indicates, the probes provide adequate spatial coverage to be used for traffic management purposes.

To examine the data coverage, the density of probes for the inner city is extracted (the results presented in the previous sections were for the entire region). Figure 6 shows the network links color-coded according to the number of probes per link for the time period 8:00 to 8:15 AM during 5 weekdays from March 1 to March 5. Green and red colors denote links with lower (<10) and higher (>=10) number of probes respectively. Gray lines are network links with zero number of probes. During that period, a total of 8,000 probes were obtained covering about 2,200 links in the (inner) network. It is clear that even for this short time period and with data from only 5 days, the probes cover the most important motorways and main roads in the inner city.

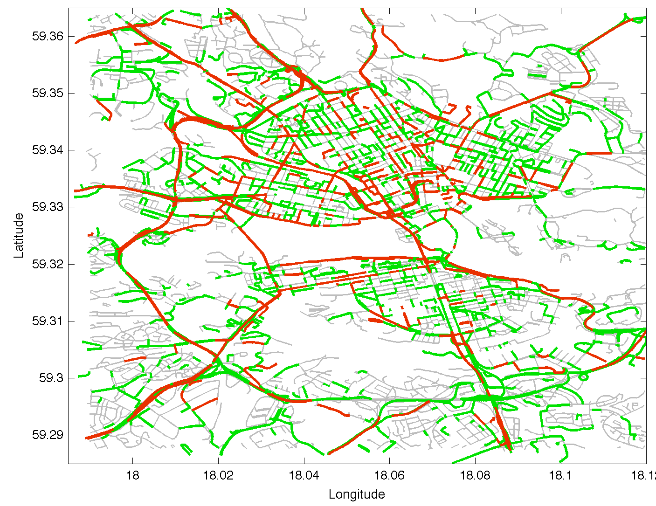

Fig. 6. Stockholm road network coverage by GPS probes. Green and red color denotes links with lower and higher number of probes respectively.

d. *Analysis of Travel Time Data*

The value of the GPS data is illustrated with an analysis of the travel time data using trips between a particular pair of origin-destination, Stockholm Central to Arlanda Airport. The distance is about 40 km, mostly a motorway (E4). FCD from taxis with customer originating from the Stockholm Central train station, and terminating at the Arlanda Airport for the 7:00-9:00 morning peak period and the 4:00 to 6:00 afternoon peak period are used for the analysis. The data cover the period between December 2009 and February 2010. The afternoon data also represents travel in the same direction.

An important attribute of GPS data is that they are almost continuous. Therefore, there is potential to collect travel time data that describe not only average values, but also their distribution. Therefore, GPS data can provide information on the variability of travel times for a specific OD pair and even more importantly time period. Since travel time reliability is an important measure of network performance, this information can be very valuable.

Figure 7 shows the variability of the travel times by day of the week and time period. The test for equality of the means shows that the null hypothesis of equal means cannot be rejected for weekdays at the 5% significance level. As expected, the average and standard deviation (std) is less during weekends. Morning and evening peak periods also exhibit different behavior. For example, the average/std of travel time in the afternoon is increased by few minutes compared to average/std of travel time in the evening.

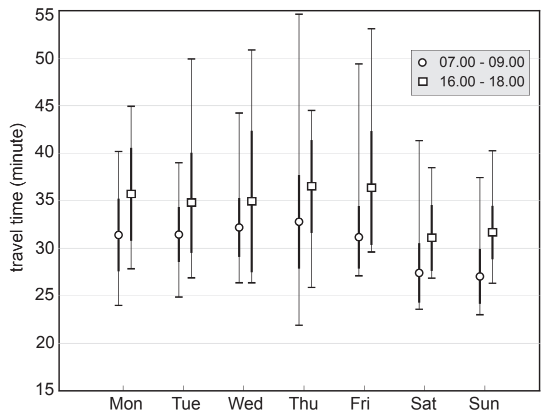

Fig. 7. Travel time variability for the Stockholm-Arlanda route. Mean (circle/box), standard deviation (bold), and min/max (bar).

The figure also illustrates quite a range of extreme travel times for the various days with significant differences between minimum and maximum observed values. In this example, it was also found that holidays produced the same patterns as weekends in terms of travel time between Stockholm central station and Arlanda airport.

IV. Conclusion

The availability of FCD as well as other data provided by emerging sensors, facilitates a number of interesting applications related to traffic monitoring and management. The paper presented the main components of an experimental ITS laboratory that is designed to provide the functionality needed to support the use of such data. The laboratory uses the stream computing paradigm to provide computational resources needed for real-time processing of the large amounts of data. The paper also presented preliminary results that support conclusions from earlier studies that GPS data have the potential, when processed correctly, to provide a wealth of unique information for traffic management, including traffic information generation.

Acknowledgments

The authors would like to thank Inga-Maj Eriksson from the Swedish Transport Administration and T. Julner from Trafik Stockholm for their support, feedback, and provision of the data for this study and Erling Weibust of IBM for facilitating the IBM/KTH collaboration.

References

[1] G. Leduc, "Road Traffic Data: Collection Methods and Applications," *Institute for Prospective Technological Studies, European Commission*, 2008.

[2] EU Galileo: http://ec.europa.eu/enterprise/policies/space/galileo accessed April 13, 2010

[3] H. Bargera, "Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: A case study from Israel," *Transportation Research Part C: Emerging Technologies*, vol. 15, 2007, pp. 380-391.

[4] S. Bekhor, M. Hirsh, S. Nimre, and I. Feldman, "Identifying spatial and temporal congestion characteristics using mobile phone data," *Proceedings of the 87th Annual Meeting of the Transportation Research Board*, Washington, D.C., 2008.

[5] D. Work and A. Bayen, "Impacts of the Mobile Internet on Transportation Cyberphysical Systems: Traffic Monitoring using Smartphon*es," National Workshop for Research on High-Confidence Transportation Cyber-Physical Systems: Automotive, Aviation, \& Rail, Washington, DC*, 2008, pp. 18-20.

[6] P. Furth, B. Hemily, T. Muller, and J. Strathman, "Uses of Archive*d AVL-APC Data to Improve Transit Performance and Management: Review and Potential," TCRP Web Document, vol. 23, 2003.*

[7] *W. Huber, M. Lä*dke, and R. Ogger, "Extended floating-car data for the acquisition of traffic information," *Proceedings of the 6th World Congress on Intelligent Transport Systems*, 1999, pp. 1-9.

[8] C. Demir, B.S. Kerner, R.G. Herrtwich, S.L. Klenov, H. Rehborn, M. Aleksic, T. Reigber, M. Schwab, and A. Haug. "FCD for urban areas: method and analysis of practical realisations," In *Proceedings of the 10th World Congress and Exhibition on Intelligent Transport Systems and Services (ITSS03)*, Madrid, Spain, November 2003.

[9] US Department of Transportation. Intelligent transport services benefits, costs and lessons learned databases, http://www.itscosts.its.dot.gov/its/benecost.nsf, 2005.

[10] M. Stonebraker, U. Çetintemel, and S. Zdonik, "The 8 requirements of real-time stream processing," *ACM SIGMOD Record*, vol. 34, 2005, p. 47.

[11] IBM InfoSphere Streams, [online] http://www.ibm.com/software/data/infosphere/streams, accessed: April 12, 2010.

[12] B. Gedik, H. Andrade, K. Wu, P. Yu, and M. Doo, "SPADE: The System S declarative stream processing engine," *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, ACM, 2008, p. 1123–1134.

[13] J. Wolf, N. Bansal, K. Hildrum, S. Parekh, D. Rajan, R. Wagle, K. Wu, and L. Fleischer, "SODA: an optimizing scheduler for large-scale stream-based distributed computer systems," *Middleware 2008*, 2008, p. 306–325.

[14] A. Biem, E. Bouillet, H. Feng, H. Koutsopoulos, C. Moran, A. Ranganathan, A. Riabov, and O. Verscheure, "IBM InfoSphere Streams for Scalable, Real-Time, Intelligent Transportation Services," SIGMOD '10: *Proceedings of the ACM International Conference on Management of Data* (to appear), 2010

[15] OPTIS consortium, G. Lind, OPTIS – Optimized traffic in Sweden, European Commission, Directorate General Energy And Transport. [online] http://tinyurl.com/optis, 2006, last access April 12, 2010.

[16] R. Schäfer, K. Thiessenhusen, and P. Wagner, "A Traffic Information System By Means Of Real-time Floating-Car Data," *Proceedings of the ITS*, vol. m, 2002, pp. 1-8.

[17] E. Brockfeld, S. Lorkowski, P. Mieth, and P. Wagner, "Benefits and limits of recent floating car data technology–an evaluation study," *for presentation at the 11th WCTR Conference, Berkeley, USA*, 2007, p. 24–28.

[18] K. Liu, T. Yamamoto, and T. Morikawa, "Feasibility of Using Taxi Dispatch System as Probes for Collecting Traffic Information," *Journal of Intelligent Transportation Systems*, vol. 13, 2009, pp. 16-27.

[19] Y. Lou, C. Zhang, Y. Zheng, and X. Xie, "Map-Matching for Low-Sampling-Rate GPS Trajectories," *research.microsoft.com*, 2009.

[20] M. Quddus, W. Ochieng, and R. Noland, "Current map-matching algorithms for transport applications: State-of-the art and future research directions," *Transportation Research Part C: Emerging Technologies*, vol. 15, 2007, pp. 312-328.