MOBILE MILLENNIUM STOCKHOLM

Andreas Allström*[a,b], Jeffery Archer[b], Alex Bayen[c], Sebastian Blandin[c], David Gundlegård[a], Haris Koutsopoulos[d], Jan Lundgren[a], Mahmood Rahmani[d], Olli-Pekka Tossavainen[e]

[a] *Department of Science and Technology, Linkoping University, Sweden*
[b] *Sweco Infrastructure, Sweden*
[c] *Department of Civil and Environmental Engineering, University of California, Berkeley, United States*
[d] *Division of Traffic and Logistics, Royal Institute of Technology(KTH), Sweden*
[e] *NAVTEQ LLC, United States*

ABSTRACT

The need for accurate real-time traffic information is growing in almost all big cities around the world. One of these cities is Stockholm, recently named as the fifth most congested city in Western Europe. The *Mobile Millennium Stockholm* project was initiated by the Swedish Transport Administration in order to address the need for a useful and timely traffic information system. The purpose of the project is to assimilate the knowledge gained from the *Mobile Millennium* project at University of California, Berkeley and develop new methods for data fusion, one of the most challenging research areas in the transport community today. The data fusion methods will utilize the potential of each data source in order to improve estimations and predictions of the traffic state. The project is a collaboration between the Swedish organizations Linköping University, the Royal Institute of Technology and Sweco Infrastructure, and UC Berkeley in the United States.

This paper describes the system architecture of the *Mobile Millennium* system developed in Berkeley and how it has been extended and adapted to estimate travel times in Stockholm. The flow-based traffic model used for estimating traffic conditions on highways, and the machine learning approach used for arterials, is also presented. Furthermore, the data sources available in Stockholm and their characteristics are described together with the work that has been done so far within *Mobile Millennium Stockholm*. Future work will mainly concern the further development of data fusion models and methods in order to allow for the introduction of new data sources.

***Keywords**: Travel time estimation, Traffic state estimation, Data fusion, Traffic information, Traffic data collection*

INTRODUCTION

As a consequence of the increased congestion in major cities, the need for more accurate traffic information with better coverage is growing all over the world. In parallel with this demand, considerably more traffic data is being collected from a variety of new and existing sources. Technologies like Bluetooth and WiFi together with different GPS-devices, smart phones, cell phone data and more traditional data sources like radar and loop detectors have the potential to create a highly comprehensive traffic database. Despite the abundance of traffic data, the types and formats available do not represent an immediate mean for improving overall data quality. Their temporal and spatial resolution as well as the aggregation, accuracy and precision differ substantially. Therefore, models and algorithms that enable an optimal combination of data are necessary in order to produce accurate and reliable estimates and predictions of the current and future traffic state.

Data fusion is an established research topic within the military and aircraft sector. [1, 2] More recently, the transport community has invested a significant deal of effort into data fusion [3, 4]. However, as newer and more cost-effective data sources are made available, the data fusion task has become increasingly more complex and consequently further innovative research is required.

Stockholm was recently named the fifth most congested city in Western Europe and the local and national road authorities must therefore work hard to resolve the ensuing congestion problems. One important part of this work is to provide accurate estimations of the current and future traffic state for a large part of the network. In order to achieve these goals, new cost-effective methods for collecting traffic data together with effective models that can fuse different types of data together need to be developed and implemented. The Swedish Transport Administration aims to create a system that can produce travel time estimations and predictions in real-time for the three major cities in Sweden, Stockholm, Gothenburg and Malmö. As a part of this ongoing work the *Mobile Millennium Stockholm* project was initiated in 2010. This project is a collaboration between the *Swedish Transport Authorities*, the Swedish organizations *Linköping University*, *the Royal Institute of Technology in Stockholm and Sweco Infrastructure*, and the *University of California, Berkeley* in the United States. The purpose of the project is to assimilate the knowledge gained from the UC Berkeley *Mobile Millennium* project and develop new methods for data fusion that can facilitate an adaptation of the system that meets Swedish requirements. The data fusion methods will utilize the potential of each data source in order to improve estimations and predictions of traffic state. The focus during the first phase of the project will be on estimating travel times for seven specific commuter routes in Stockholm.

The paper is structured as follows. First the *Mobile Millennium* system developed in Berkeley is described. The system architecture, together with the flow-based traffic model used for estimating traffic conditions on highways and the machine learning approach used for arterials, is presented. Next the data sources currently available in Stockholm and their characteristics, together with the work that has been done so far within *Mobile Millennium Stockholm*, is described. Finally the future work, which mainly consists of developing new data fusion methods, is presented.

Corresponding author – Linköping University, Dept of Science and Technology, SE-601 74 Norrköping, Sweden, Tel: +46 (0)11 36 32 91, Fax: +46 (0)11 36 32 70, E-mail: andreas.allstrom@liu.se

MOBILE MILLENNIUM

Mobile Millennium is a research project developed at the University of California, Berkeley. It focuses on the design, implementation, and operational deployment of novel algorithms and innovative techniques to address current road traffic challenges. The project [5] was launched jointly with Nokia and NAVTEQ, under the umbrella of the California and US Department of Transportation in November 2008. This followed the *Mobile Century* experiment [6, 7], which successfully demonstrated on a large scale the added value of user-generated GPS probe data. Hosted by the *California Center for Innovative Transportation*, *Mobile Millennium* currently collects more than 60 million GPS traffic data points daily in addition to numerous other static data feeds. The project represents state-of-the-art research on traffic modeling and estimation, routing directions and pollution estimation.

System architecture

The *Mobile Millennium* system architecture is organized around the concept of a *model graph* which consists of a graphical abstraction of the road network together with a minimal set of information required for traffic estimation. This representation is shared by all processes running in the system and practically serves as a geographic information system layer dedicated to statistical and model-based traffic estimation. The simplicity of this representation and the standardized methods developed for its generation from classical digital maps guarantee fast portability of the system to other locations. The flow of information through the system and its different modules are outlined in Figure 1.

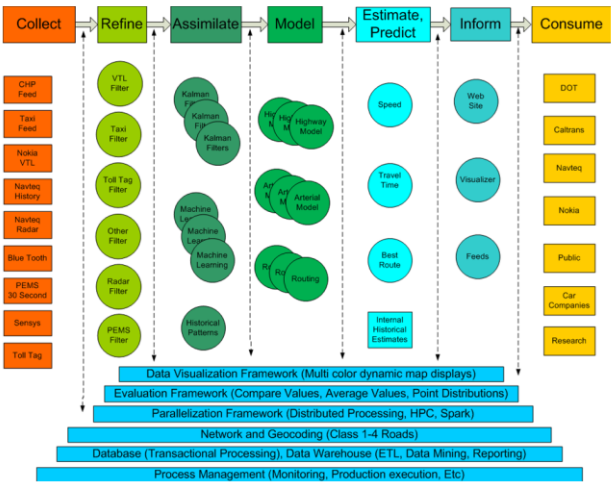

Figure 1 *Mobile Millennium*: A variety of data feeds from different data sources is processed and fed in real-time to statistical and model-based estimation algorithms. The real-time output of these algorithms is then broadcast to different classes of clients under the form of link travel-time distributions, velocity maps, and optimal routing policies.

Measurements collected by fixed traffic sensors (loops, radars, toll transponders), and commuters (GPS point speeds, GPS travel-times) are filtered in real-time and map-matched to the model graph. These traffic measurements are assimilated by physical and statistical models of traffic, and optimally combined with historical traffic knowledge to produce the best estimates of the current traffic conditions. These estimates can be broadcast to different clients or fed to routing algorithms providing commuters with optimal routing directions given the current state of the road network.

Different traffic dynamics properties and different data types and data volumes on highway and arterial networks motivate different algorithmic approaches for traffic estimation on each of network type.

Highway traffic

The *Mobile Millennium* highway traffic estimation model [3] represents the evolution of highway velocities in the San Francisco Bay Area according to the mass conservation equation, the Lighthill Whitham Richards Partial Differential Equation. Traffic state is discretized into cells of approximately 300 meters on the whole San Francisco Bay Area, and updated every 30 seconds. At every time step, the discretized partial differential equation is solved numerically using the Godunov scheme [8] on the 6000 cells composing the network for northern Californian highways.

Available measurements coming from loop detectors, radars and probe data are assimilated into the model after filtering and map-matching. The measurements are modeled as linear observations of the current traffic state and the conditional distribution of the traffic state given the measurements is computed using the ensemble Kalman filter [9]. The output of the filter is the minimal variance estimate given the traffic model and the measurements.

Arterial traffic

The *Mobile Millennium* arterial traffic estimation module consists of a Bayesian network model of link travel-times [10] in San Francisco. The statistical model represents the spatial conditional distributions of individual travel-times represented as random variables. At every time step, an *expectation-maximization* algorithm updates the distribution parameters based on the new measurements and thereafter the most likely traffic conditions are computed.

For San Francisco, the arterial model mostly relies on data from a fleet of 500 taxis that reports GPS position data every minute. One of the challenges with this type of data is related to its inherent sparsity and measurement noise, which does not allow the taxi traces to be uniquely map-matched to the model graph. To solve this problem, a dedicated *hidden Markov model* map-matching algorithm has been developed [11], which computes the probability of all possible paths between consecutive GPS points from the same taxi and produces the most likely mapping of the taxi traces. This can then be integrated into the Bayesian link travel-time model of traffic.

CURRENT STATUS IN STOCKHOLM

The traffic data collected in Stockholm is in many ways similar to the data collected in the San Francisco Bay Area. Fixed detectors and probe vehicles, mainly taxis, are the main data sources. The characteristics of these data sets are described below.

Fixed detectors

There are around 1500 fixed detectors in and around Stockholm, these are primarily located on highways. Around 1000 of these detectors are radar detectors that are an integral part of a Motorway Control System on the main highway that passes through Stockholm. These radar detectors are situated approximately every 500 meters and record data specific to each traffic lane. They collect speed and traffic flow data that is aggregated over a one-minute period. The remaining 500 detectors are a combination of loop detectors, cameras and other radar detectors. Some of these can record data for each vehicle passage while others report aggregated data for each one minute period.

The investment and maintenance cost for the fixed detectors is considered to be very high and therefore the Swedish Transport Authorities are trying to find ways to reduce them in number. Another problem with the radar detectors is that they have difficulties detecting vehicles when passage speeds are low during periods of heavy congestion. One possible way to solve these problems is to use advanced traffic models and data fusion algorithms and combine the data from fixed detectors with data from other sources, such as probe vehicles. These issues will be an essential part of this project.

Probe data

There are around 1500 probe taxis operating in the Stockholm area. The data reported includes time-stamped location (latitude and longitude) and status (free or meter-on). The taxi data is anonymized. The total number of observations per month is approximately 10 million. On average the data is already 39 seconds old by the time it arrives in the database. The transmission frequency depends on the status of each taxi (occupied, free, etc.), and is approximately 110 seconds on average. The total number of probes reported during any particular period of time depends on the number of taxis that were active and their respective reporting frequency. The two identified traffic peaks, morning and early evening during week days, represent periods when there is most taxi demand in the city; hence a larger number of probes and taxi data is reported during these peak periods. To examine the data coverage, the density of probes for the inner city has been extracted. Figure 2 shows network links that have been color-coded according to the number of probes per link during the morning period 08:00-08:15 on five weekdays between March 1 and March 5, 2010. The observations cover the majority of the links of interest in this project and facilitate the development of a historical database that is critical for a number of applications.

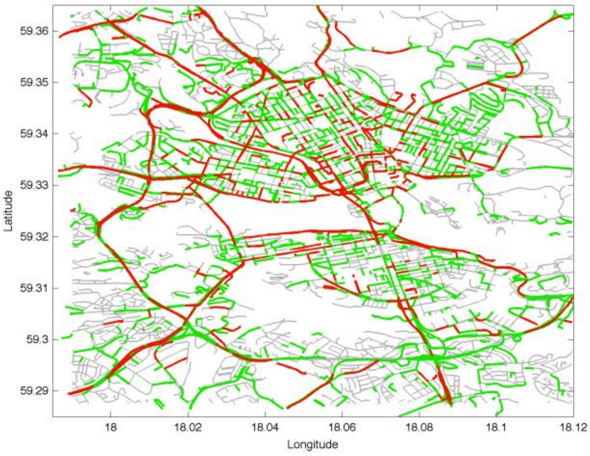

Figure 2 Network links color-coded according to the number of taxi probes per link for the time period 8:00 to 8:15 AM during five weekdays from March 1 to March 5 2010. Green and red colors denote links with lower (<10) and higher (>=10) number of probes respectively. Gray lines are network links with zero number of probes.

In addition to the taxi probes, around 150 distribution vehicles also act as probe vehicles and report their position, speed and direction every 30 seconds.

LPR-cameras

Travel times are collected for approximately 100 routes in and around Stockholm. Travel time data is collected through a system based on license plate recognition cameras. These are installed at locations on important routes all over the city. Travel time data is reported for each individual vehicle. A problem with this data is that the travel time data is reported after passing the camera at the end of a measured route. Thus, the data is already old making it also hard to detect congestion situations in real-time. Nevertheless, this data is extremely useful for validation purposes and can potentially be used for more real-time settings when a historical database has been created.

Data processing today

The data processing carried out in the current travel time system implemented at the Traffic Management Centre in Stockholm is relatively unsophisticated. A simple data fusion algorithm is applied for links where travel times are calculated from two or more data sources. In the algorithm, a quality value allocated to each measurement is used for weighting purposes. For links where no real-time data is available the travel time is calculated from historical data as well as data from neighboring links. Currently, there are no predictions of travel time in Stockholm.

Besides the real-time calculations of travel time, a travel time calculation for applications that do not work in real-time is also made. Examples of such applications include a historical database with indicators that describe the congestion level in Stockholm for the long–term. The historical database contains all traffic data collected since 2004.

Mobile Millennium Stockholm

During 2010 the *Mobile Millennium* system was adapted to handle traffic data collected in Stockholm and to estimate the traffic state on highways and arterials in Stockholm. As mentioned earlier, the data collected in Stockholm is similar to that collected in the San Francisco Bay Area. There are, however, many subtle differences in the characteristics of the data which require adjustments of the filters in the system. The model graph which forms the foundation of the *Mobile Millennium* system architecture has been created for the Stockholm implementation.

During 2010 a small field trial involving ten probe vehicles with GPS-equipped cell phones was carried out in Stockholm. The data collected in this field trial play an important part in the validation and calibration of the implemented *Mobile Millennium* system. Furthermore, a simulation model that can reproduce the data collected on the main highway in Stockholm is being built.

FUTURE WORK

The *Mobile Millennium* system represents a great potential for improving traffic state estimation and prediction in the Stockholm area and in other Swedish cities. However, since the data and the behavior of the drivers are somewhat different to that of the corresponding Berkeley project, the models require adaptation and redevelopment as well as careful validation and calibration. Furthermore, work with the data fusion algorithms must allow for the introduction of new data sources that can improve the overall quality of traffic information.

Highway model

The work with the highway model is currently focused on how to fuse spatially dynamic section speeds from taxi probe vehicles with point speeds from stationary sensors. The *Mobile Millennium* project successfully managed to fuse measurements from stationary sensors with measurements from probe vehicles. [12] The probe vehicles in that field test used a spatially uniform sampling known as Virtual Trip Lines and the data was fused using the Cell Transmission Model for velocity (CTM-v) as traffic model and ensemble Kalman filtering for assimilation. This effort will be continued, in this case focusing on the types of probe vehicles available in Stockholm, i.e. the aforementioned taxis, distribution vehicles and eventually GPS-equipped cell phones. The data fusion has a great potential of improving the accuracy of the travel time estimation, especially during congestion, and possibly reduce the number of fixed sensors on the highway.

Arterial model

The taxi GPS data provide useful information regarding traffic conditions on arterials in the Stockholm region. Future work includes the development of the historical database about speeds on links in the network and their distribution, as well as testing of different fusion algorithms for traffic estimation and prediction.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the systems contribution of Saneesh Apte and Daniel Edwards, without whom this project would not have been possible, and the careful guidance of the *Mobile Millennium* project manager Joe Butler.

REFERENCES

[1] Gustafsson F.,(2010). *Statistical Sensor Fusion*, Studentlitteratur, Lund, Sweden.
[2] Hall D.L., Llinas J. (1997). An Introduction to Multisensor Data Fusion, *Proceedings of the IEEE*, Vol. **85**, pp. 6-23, Jan. 1997
[3] Work, D., Blandin, S., Tossavainen, O.-P., Piccoli, B., and Bayen, A. (2010) A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress* **2010(1)**, 1-35.
[4] van Lint J.W.C. , Hoogendoorn S.P.. (2010). A Robust and Efficient Method for Fusing Heterogeneous Data from Traffic Sensors on Freeways, *Computer-Aided Civil and Infrastructure Engineering* **25** (2010) 596–612
[5] Mobile Millennium http://traffic.berkeley.edu.
[6] Amin S., (2008). *Mobile Century Using GPS Mobile Phones as Traffic Sensors A Field Experiment*, In Proceedings of the 15th World Congress on Intelligent Transportation Systems, New York, NY, November 16-20, 2008.
[7] Herrera, J.-C., Work, D., Ban, X., Herring, R., Jacobson, Q., and Bayen, A. (2009). Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment, *Transportation Research Part C,* **18**, 568-583.
[8]Lebacque, J. (1996) The Godunov scheme and what it means for first order macroscopic traffic flow models, *Proceedings of the 13th ISTTT, Ed. J.B. Lesort*, pp. 647-677.
[9]Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean dynamics* **53(4)**, 343-367.
[10]Herring, R., Hofleitner, A., Abbeel, P., and Bayen, A. (2010), Estimating arterial traffic conditions using sparse probe data, In 13th International IEEE Conference on Intelligent Transportation Systems, Madeira Island, Portugal, Sep. 19-22, 2010.
[11]Hunter, T., Herring, R., Abbeel, P., and Bayen, A. (2009) Path and travel time inference from GPS probe vehicle data, In Neural Information Processing Systems foundation (NIPS), Vancouver, Canada, December ,2009
[12]Claudel C.G.., Hofleitner A., Mignerey N.D, Bayen A. (2009). 88th Transportation Research Board Annual Meeting, Washington D.C., January 10-14 2009