The measurement architecture of the Virtual Traffic Laboratory

a design of a study in a data driven traffic analysis

Arnoud Visser, Joost Zoetebier, Hakan Yakali and Bob Hertzberger

Informatics Institute, University of Amsterdam

Abstract. In this paper we introduce the measurement architecuture of an application for the Virtual Traffic Laboratory. We have seamlessly integrated the analyses of aggregated information from simulation and measurements in a Matlab environment in which one can concentrate on finding the dependencies of the different parameters, select subsets in the measurements, and extrapolate the measurements via simulation. Available aggregated information is directly displayed and new aggregate information, produced in the background, is displayed as soon as it is available.

1 Introduction

Our ability to regulate and manage the traffic on our road-infrastructure, essential for the economic welfare of a country, relies on an accurate understanding of the dynamics of such system. Recent studies have shown very complex structures in the traffic flow [1]. This state is called the synchronized state, which has to be distinguished from a free flowing and congested state. The difficulty to understand the dynamics originates from the difficulty to relate the observed dynamics in speed and density to the underlying dynamics of the drivers behaviors, and the changes therein as function of the circumstances and driver motivation [2].

Simulations play an essential role in evaluating different aspects of the dynamics of traffic systems. As in most application areas, the available computing power is the determining factor with respect to the level of detail that can be simulated [3] and, consequently, lack of it leads to more abstract models [4]. To be able to afford more detailed situations, we looked how we could use the resources provided by for instance Condor[5], or the Grid [6].

Simulation and real world experimentation both generate huge amount of data. Much of the effort in the computer sciences groups is directed into giving scientists smooth access to storage and visualization resources; the so called middle-ware on top of the grid-technology. Yet, for a scientist seamless integration of the information from simulated data and measurements is the most important issue, the so called data-driven approach (see for instance [7]).

In this article we show our experience with building our Virtual Traffic Laboratory as a data driven experimentation environment. This experience can be used as input for the future development of the Virtual Laboratory on other application domains.

2 Measurement Analysis Architecture

Traffic flow on the Dutch highway A12 is investigated for a wide variety of circumstances in the years 1999-2001. This location has the unique characteristic that, although the flow of traffic is high, traffic jams are very sporadically seen. In this sense it is a unique measurement point to gather experimental facts to understand the microscopic structures in synchronized traffic states ([1]), which was not reported outside of Germany yet.

For the understanding the microscopic structures in synchronized traffic states the relations between several aggregates of single vehicle measurements have to be made. Important aggregate measurements are for instance average speed, average flow, average density, headway distribution and speed difference distribution. The dynamics of these one-minute aggregates over 5-10 minutes periods are important for a correct identification of the state.

To facilitate the analysis of aggregate measurements over time we designed the following architecture:

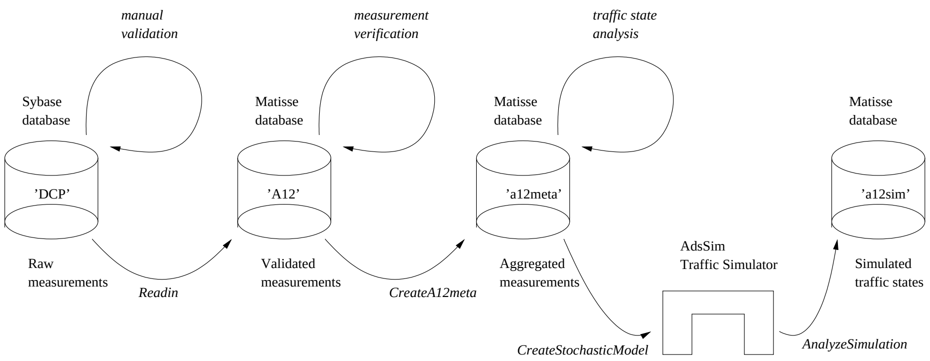


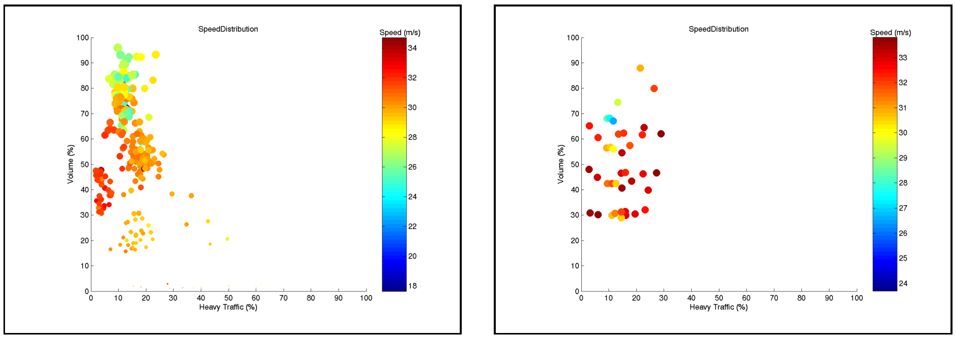
Fig. 1. The measurement analysis architecture

Along the A12 there was a relational database from Sybase that collected the measurements from two independent measurement systems. One system was based on inductive loops in the road, the other on an optical system on a gantry above the road. Although both were quality systems, some discrepancies occur between measurements due to different physical principles. Video recordings were used to manually decide the ground truth when the measurements were not clear.

After this validation process, the measurements were converted to an object oriented database from Matisse. This database was used to verify the quality of the measurement systems themselves. While the manual validation process was used to get the overall statistics of errors in the measurements, the object oriented database was used to analyze the circumstances of the measurement errors.

The validated measurements were used to generate the statistics that characterize the traffic flow. Different measurements-periods could be combined based on different criteria. The right combination of criteria results in candidate traffic flow states. The statistics that are important to characterize the microscopic structure of the traffic flow require fits of complex (non-Gaussian) probability density functions. The statistics were stored as meta-data in a separate database.

An example of such analysis is given in figure 2, where the average speed is given as a function of the flow (as percentage of the maximum flow) and the fraction of lorries (as percentage of the number of passages).



(a) measured

(b) simulated

Fig. 2. The average speed as function of the flow and the fraction heavy traffic

The average speed is indicated with a colorcode, red (top of the bar) indicates high speeds, blue (bottom of the bar) indicates low speeds. Each point indicates an aggregate over longer period (30-60 minutes), which are typically equivalent with a few thousand passages.

Combinations of measurement-periods that showed the same patterns on their aggregated traffic flow measurements over time were candidate traffic flow states. These aggregated measurements could be translated into the parameters of a microscopic traffic simulator, AdsSim [8], which is based on the microscopic Mixic model [9].

The characteristics of the simulated data were aggregated in the same way as the real data, and the resulting dynamics were compared to the original dynamics, to see if the model was complete (see figure 2). As one can see, the simulated points are more homogeneous spread over the spectrum because one can generate a dependency grid. Yet, the results are less to be trusted when one has to extrapolate far from actual measured parameter-combinations space.

The *CreateA12meta* and *CreateStochasticModel* were originally Matlab-functions with more than 1000 lines of code. We converted those functions to standalone programs, which made it possible to run those functions in the background with the aid of the Condor software[5].

3 Discussion

We have chosen this application, because of the complexity of both the measurement analysis and the traffic flow model. For instance, the Mixic model has 68 parameters in its traffic flow model [9], and most parameters are described as functions of single vehicle data such as lane, speed and headway. For AdsSim this resulted in 585 variables that can be adjusted to a specific traffic condition. Compare this with the 150 keywords in the standard application in molecular dynamics [10] in the UniCore environment [11].

To be able to calibrate such a model for a certain traffic state, one needs to be able to select characteristic subsets in the bulk of measurements, and visualize the dynamics of the aggregates in different ways. It is no problem that it takes some time to generate aggregates, as long as one is able to switch fast between diagrams of parameters and their dependencies as soon as the aggregates are ready. Storing the analysis results in a database solves this problem.

When a certain pattern in the dependency is discovered, one can select measurements and simulation to add extra points to the diagram, to look if the pattern holds. While processes at the background fill in the missing data-points, the Scientist starts the visualization of other dependencies, till an unexpected pattern appears. At that moment other subsets in the measurements become important. This means that new analysis has started up, and the decision has to be made to stop the current analysis.

This approach differs from the approach of for instance [7], where Matlab is seen as inappropriate due to license policies and speed issues. By using high throughput systems in the background speed is no longer an issue. Further both approaches do not exclude each other. In our view the programming effort can concentrate on often used algorithms, and optimize these algorithms into modules that can be executed in the background, while Matlab is used for direct analysis and visualization.

4 Conclusions

In this article we have introduced the measurement architecture of an experiment in the Virtual Traffic Laboratory. To aid the scientist, analysis results are stored in databases with aggregated data. This allows to repeatedly display the results from different viewpoints, where the scientist does not have to worry that too rigorous filtering will force him to do the aggregation again.

New aggregate data can be generated by exploring a dependency by performing new analysis on sets selected on different parameter-combinations in the background. This analysis can be performed seamlessly on both real data and simulated data. New data can be automatically displayed by adding monitors to the databases.

References

1. L. Neubert, et al., "Single-vehicle data of highway traffic: A statistical analysis", Physical Review E, Vol. 60, No. 6, December 1999.
2. Th. Jörgensohn et al., "Modelling Human Behaviour, a Must for Human Centred Automation in Transport Systems?", Proc. BASYS 2000, Berlin, Germany, September 27-29, 2000.
3. K. Nagel, M.Rickert, "Dynamic traffic assignment on parallel computers in TRANSIMS", in: Future Generation Computer Systems, vol. 17, 2001, pp.637-648.
4. A. Visser et al. "An hierarchical view on modelling the reliability of a DSRC-link for ETC applications", IEEE Transactions on Intelligent Transportation Systems, Vol. 3: No. 2, June 2002.
5. Douglas Thain, et al., "Condor and the Grid", in Grid Computing: Making The Global Infrastructure a Reality, John Wiley, 2003. ISBN: 0-470-85319-0
6. I. Foster, C. Kesselman, "The Grid: Blueprint for a New Computing Infrastructure", Morgan Kaufmann, 1999.
7. Craig C. Douglas, "Virtual Telemetry for Dynamic Data-Driven Application Simulations"; Proc. ICCS 2003, Melbourne, LNCS 2657. p. 279-288, Springer-Verlag 2003.
8. A. Visser et al. "Calibration of a traffic generator for high-density traffic, using the data collected during a road pricing project", paper 4052 to the 9th World congress on Intelligent Transport Systems, Chicago, Illinois, October 2002
9. C. Tampére, C. "A Random Traffic Generator for Microscopic Simulation", Proceedings 78th TRB Annual Meeting, Jan. 1999, Washington DC, USA.
10. W. Andreoni and A. Curioni "New Advances in Chemistry and Material Science with CPMD and Parallel Computing", Parallel Computing 26, p. 819, 2000 .
11. D.W. Erwin et al. "UNICORE: A Grid Computing Environment", in LNCS 2150, p. 825-839, Springer-Verlag Berlin Heidelberg, 2001.