Data Management for Intelligent Transport System Using Pervasive Sensing

V Suresh¹, P Watson¹, J Neasham², M Bell³, Dave Pearson⁴,

David Oliver⁵, Fabio Galatioto³, Graeme Hill³, Jay Parmar⁶

1 School of Computing Science, Newcastle University, NE1 7RU, <u>visalakshmi.suresh@ncl.ac.uk</u>, 2 School of Electrical and Computing, Newcastle University, 3 TORG, School of Civil Engineering and Geosciences, Newcastle University, 4 Vice President, Architecture, Global Technology Business Unit, Oracle Corporation, 5 Newcastle City Council, Traffic Signal group,6 Leicester City Council

ABSTRACT

In this paper we address the information management framework to deliver the design and implementation of a static and mobile, detector system for monitoring traffic and pollution in existing road networks. The measured data comes from networks of low cost, pervasive sensors referred to as motes or smart dust, as well as legacy monitoring stations installed by city councils to monitor traffic and pollution. The framework provides real-time information on traffic, pollution and meteorological conditions. The streaming data from the heterogeneous data sources is collected, processed, analyzed and disseminated so that it can be used to drive applications in real-time and collect historic information in a data warehouse. The real time and historic data is integrated to monitor network status, manage traffic to reduce congestion, improve air quality and manage noise impacts.

1. INTRODUCTION

<u>Mobile Environmental Sensing Across Grid (MESSAGE)</u> [2] is a project jointly funded by the EPSRC and Department for transport, UK, aims to develop an environmental pervasive sensor system for urban road networks. Newcastle University has developed wireless sensors which are fixed in the street infrastructure, carried by vehicles and pedestrians thereby collecting data relating to the pollution and traffic at a very high spatial resolution. The collected data is sent, using wireless GPRS links in real-time, to the central server.

Our architecture is based on a central repository (section 2) to integrate real time streaming data from sensors, traffic detectors and consolidate the view across various data sources to support real time data dissemination decision-making at various levels such as general public, policy makers and local authorities. We standardize the data from the source systems, produce historic trend analysis and compare the traffic and pollution profile over the period of time in the data warehouse (section 3.2). We adopt the "bottom-up" approach by building separate data marts for critical subject areas namely pollution and traffic flow. The data in the real time and historic database forms as a basis to monitor the status of the network in the short term as well as daily, weekly, monthly and annual average concentrations for longer term environmental assessments using a user friendly interface for decision makers.

To date, many research activities have been undertaken to build transport warehouse for the collection of traffic data. However, substantial analysis of this data has to be achieved offline and it is rare for historic and real time data to be integrated with models to create an intelligent decision support system. This work seeks to develop an online information management system for measuring, managing and mapping traffic and pollution (section 3.3).

2. Framework

The overall aim is to provide a general data management framework for supporting the real-time and historic analysis of transport data, moving away from current ad-hoc approaches. We introduce a novel prototype system that is currently deployed in Gateshead, Leicester, Newcastle, London and Palermo (Italy). Consequently, it has been designed to handle very high data input rates from hundreds of sensors, along with the analysis of very large amounts of historic data (collected over 20 years) in the warehouse. Data from the real time database adheres to UTMC [1] standards where the data is cleaned, filtered and transformed in to a Data Warehouse. The data is used by applications and users to explore historic trends and derive predictive mathematical and micro simulation models. These models can then be used by the real time applications to inform decisions taken to reduce congestion and pollution. The framework uses the Oracle database 11g [8], Oracle sensor edge server and Oracle Business Intelligence tools are customized for the reporting. The complete framework is shown in figure 1. Along with the sensor network, other streaming events are collected from data sources such as inductive loop traffic detectors. Data management is categorized in three different aspects namely real-time monitoring, historic data analysis and integration of real-time events with historic analysis for decision making. To handle the linearly increasing data volume, data warehouse with dimensional design is implemented. The concepts of star transformations, various techniques of partitioning, indexing and materialized views are being used to achieve the desirable query performance. With reference to Intelligent Transport Systems, various levels of aggregations are carried out to slice and dice the data for knowledge discovery and to meet the needs of different application domains for example transportation, traffic micro simulation, air quality , noise models evaluation, impact assessment and so on.



Figure 1: Information Management Framework

3. Information Management:

3.1 Real Time Architecture:

The data sources such as pervasive sensors, metrological stations and traffic loop detectors use queuing systems. The design behind these systems is to propagate and consume the large amounts of data as fast as possible. Database oriented queuing systems meet the scalability and performance characteristics such as sophisticated buffer management capabilities [4]. The data from the sensors and other data sources are streamed in to the data base using oracle sensor edge server that supports common operations for adding, configuring and removing devices, filters and dispatchers. The real-time streaming data undergoes quality assurance. In addition appropriate calibration of sensors based on the temperature and humidity measured simultaneously at each location is achieved by applying business rules to classify the data based on their origin. In addition to the temperature and humidity correction, calibration [6] process corrects the data collected by the sensors in terms of the drift and manufacturers offset specified in the electrochemical hardware components. The filtered and normalized data is forwarded to the database using streams dispatcher which utilizes the thread pooling and throttling to reduce the impact on the main data streams. The archiving facility in the edge server helps to retrieve data in case of network failure or systems failure. The Oracle Streams is used in our system as illustrated in the figure 2 to capture the real time data and also publishes the data to the applications. The Streams resides on top of the change data capture providing full transport mechanism for data synchronization. It uses advanced queuing techniques to publish and subscribe events in queue.

The traffic congestion, emission and dispersion are subscribed to the streaming events to disseminate the information in maps and user friendly applications as in figure 2.





Figure 2: Real Time architecture and visualization to capture streaming events

3.2 Historic Data Warehouse

The warehouse schemas are populated by subscribing to the oracle streams. Facts about pollution and traffic are built using star schema focusing the analytical processing.



Figure 3: Pollution and Traffic Flow Dimension Model

Historic data marts as shown in Figure 3 are created for calculating the daily pollution and traffic flow profiles as shown in the Figure 4. The data marts are referred with different dimensions based on the time, location of measurement, type of the device and meteorological conditions.



Figure 4: Pollution profile for a period of 2 months monitored in Leicester

3.3 Real Time and Historic Data Integration

In a typical road network, the intersection of two roads is referred to as nodes and the road segments between these intersections, the nodes are links. Historical analysis is started at the link level for mining patterns of traffic flow. For each link, the traffic flow pattern is analyzed and relationships are derived for congested, busy and free flow of traffic for example as shown in the Figure 5. These relationships are used to create algorithms to support the decision making queries to be programmed for applications.



Figure 5: Real-time congestion status

The congestion parameter may vary for each and every link and node in the real-world scenario and the challenge here is to develop generic statistical methods to automatically create the intelligence needed to deliver the decision support. In order to monitor the regional pollution and traffic, data from thousands of links and nodes should be processed. The pattern of pollution and traffic varies according to external influences such as accidents, events including football match, road works etc. which occur at random or preplanned. The monitored pollutants, CO and NO2 are data mined for different locations and patterns are investigated to allow forecasting of future events or trends to provide timely and more effective decision support. The historic analysis of the traffic over a period of time, involves a study of pattern of traffic in each segment of the road. This pattern is recorded and compared with the real time streaming data to categorize the nature of traffic. The data is published as a data service and given access to user friendly applications. The system, tested in Newcastle, is already able to display by a trial web-service realtime traffic conditions (according to four statuses: congested, busy, smoothed flow and free flow each state associated with a different color). In addition we are developing ways to also provide an indication of link emissions (Figure 5) and eventually an estimate of the local pollution concentration as described in the workflow. The canyon pollution dispersion maps based on pollutants such as CO, NO2 and PM10 are created using OSPM [8] as indicated in the workflow as shown in Figure 6.



Figure 6: Emission and Dispersion Workflow

4. CONCLUSION

This paper presents an overview of integration and management of data from disparate data sources in the context of both realtime and historic analysis. As a basis for deriving meaningful information from the data sets for Intelligent Transport Systems which is subsequently to define and implement appropriate control and management actions. Various standards and business policies have to be applied to ensure stakeholder operating requirements. The life cycle of the data from various data sources involves real time data capture, processing, storage, analysis and dissemination. Combining real-time data from data sources such as pervasive sensors with existing legacy traffic systems will not only improve knowledge in traffic flow, congestion and pollution but should also reveal new information for more effective control and to achieve added value from Intelligent Transportation Systems implementation.

5. ACKNOWLEDGMENTS

Our thanks to EPSRC and DfT for funding the project, Gateshead City Council, Newcastle Traffic Signal Group, Leicester City council for their continued support to supply and allow installation of pervasive sensing system to deliver the richness of data needed for this research. We thank Oracle UK for their support and advice in the product customization.

6. REFERENCES

- Urban Traffic Management Control UTMC http://www.dft.gov.uk/pgr/roads/tpm/utmc/
- [2] MESSAGE Mobile Environmental Sensor System Across Grid <u>www.message-project.org</u>
- [3] The Data warehouse Toolkit Ralph Kimball
- [4] <u>Message Queuing for Business Integration</u>
- [5] Bell M.C, Suresh V., Watson P., Blythe P.T. (2008a) UTMC Compliant Database to support technologies of the future. RTIC Conference 20th 22nd May 2008 Manchester
- [6] Bell M.C., Neasham, J.A., Hill G., Tate J.E., (2008c)
 Validation of carbon monoxide and nitrogen dioxide sensors. Internal Report, Transport Operations Research Group, Newcastle University, November 2008
- [7] Bell M.C., Galatioto F., Hill G., Mohammad A., Automatic Evaluation of Environmental impacts of traffic DEMAND management Strategies (TDMS), UTSG, Jan 2009
- [8] OSPM Berkowicz, R., Hertel, O., Larsen, S.E., Sorensen, N.N., Nielsen, M., 1997. Modelling traffic pollution in streets. NERI Report, Roskilde, Denmark.