A QUALITY- AND PRIORITY-BASED TRAFFIC INFORMATION FUSION ARCHITECTURE

Sandra Geisler*, Christoph Quix*, Guido G. Gehlen**, Gordian Jodlauk**

*Information Systems, RWTH Aachen University, Ahornstraße 55, 52076 Aachen, Germany, {geisler,quix}@dbis.rwth-aachen.de

**Ericsson GmbH Eurolab, Ericsson Allee 1, 52134 Herzogenrath, Germany
{Guido.G.Gehlen,gordian.jodlauk}@ericsson.com

ABSTRACT

A novel architecture is proposed that enables the detection, automatic priority-based processing and validation of event messages, and data fusion of periodically gathered probe data to facilitate various Intelligent Transportation System (ITS) applications. Floating car data and mobile phone data are processed in a multilevel Event Validator service to identify road traffic events. Based on the data quality and the priority of the identified event, it is either sent to other cars in the area or processed further using additional information to achieve a higher reliability of the detected event.

KEYWORDS

Floating Car Data, Cellular Networks, UMTS, Data Fusion, Traffic Information System, Positioning, WLT-based technologies, Aktiv, Cooperative Cars, CoCar

INTRODUCTION

For traffic management and traffic information services, reliable and high-quality data is indispensable. Data sources delivering road traffic monitoring information nowadays comprise mainly stationary detectors, such as inductive loops, laser scanners, or camera technology. However, stationary detectors have high deployment and operational costs and hence cannot be installed over large areas of
the road network. Therefore, mobile detection solutions, namely Floating Car Data (FCD) technologies, gain more and more attention in recent transportation research, such as the Floating Car Observer approach [1]. There exist different wireless location-based technologies (WLT) which enable the acquisition of traffic data based on mobile probes. These techniques can be partitioned into handset-based (such as GPS) and network-based systems (such as CellID or Angle of Arrival) depending on the site which is in charge of the positioning [2]. Especially positioning techniques based solely on mobile phones carried in the car seem appealing as they have low deployment costs to fast market introduction. These techniques can be utilized to implement Car-to-Car or Car-to-Infrastructure architectures as they allow for immediate event detection. To generate valuable, expressive and high-quality road traffic information for Intelligent Transportation Systems the acquired data from the different data sources is integrated by Data Fusion.

The contribution of this paper is a novel architecture for data processing in traffic information systems using extended floating car data (XFCD) and mobile phone data. We employ a multi-level quality-oriented and priority-based approach to detect events and integrate data from different sources to estimate the traffic state. The input data is analyzed level by level, if the derived information has a certain quality and also depending on the priority of the message, the information is sent to system clients, such as cooperative vehicles. Otherwise, further data analysis techniques will be applied on the next level, possibly including additional data sources, to derive information with higher quality. This implies a data hierarchy in which raw data is refined and supplemented from level to level by using additional information.

The approach is based on an open architecture in which additional information sources can be integrated in a flexible way. An important component is a Data Stream Management System which allows integration and analysis of incoming streams of data. New sources can be added by defining new data streams and integrating these with the existing data streams. The application of our proposed architecture is not limited to traffic state estimation, more fine grained events such as an emergency braking can be detected and broadcasted to the vehicles in the vicinity of this event.

In contrast to other architectures, such as proposed in CVIS [16], COOPERS [17], NoW [20] or CarTel [21] (see also the Related Work section), our envisioned architecture combines several features which are supposed to ensure a fast, cost-saving and high-quality processing and utilization of traffic data from different sources for traffic state estimation and short-term forecasting. The architecture is designed to mainly use data from mobile sensors, which saves deployment costs and enables the coverage of a larger part of the road network. It further reuses the existing infrastructure of cellular networks for data acquisition (using anonymous mobile
phone data) as well as for sending and receiving event messages and probe data in the system. This additionally reduces costs, because no new infrastructure has to be provided in comparison to WiFi-based solutions. Furthermore, to enable timely data processing and delivery to applications, a Data Stream Management System is integrated into the architecture. It is responsible for data cleaning, data transformation and fusion of data. The data quality measures, event validation and prioritization mechanisms provide the possibility to dynamically adapt speed and accuracy of traffic information processing in contrast to other architectures. The more urgent an event message is, the faster it will be processed. But the longer a message stays in the system, the more accurate the information of the message will get, as data from more and more vehicles, phones and from a bigger time interval can be used to verify the event. The architecture is processing and using floating vehicle and phone data as well as event messages.

The context of this work is the Cooperative Cars project (CoCar). The CoCar project, funded by the German Federal Ministry for Research and Education, investigates the suitability of UMTS technologies and their foreseeable extensions for direct, targeted transmission of traffic data. The project is a part of the research initiative “Adaptive and Cooperative Technologies for the Intelligent Traffic” (AKTIV). In the project it is identified which traffic management and driver assistance applications are suitable for use of the aforementioned technology [3] [22].

In the following, we first describe our envisioned priority- and quality-based data fusion architecture, which implements the data processing for the Cooperative Car applications and its most important components. Thereafter, the fusion process of the different data sources enabled by a Data Stream Management System is described in more detail. Furthermore, the data quality and prioritization aspects of the architecture are illuminated. Finally, related work from the fields of Car-2-X architectures and Data Stream Management is presented.

ARCHITECTURE

As mentioned before, data from different sources can enhance the quality and reliability of road traffic information. The envisioned architecture, as depicted in Figure 1, comprises different data sources which help to detect a certain traffic situations or incidents. Predominantly, data is gathered inside the CoCar vehicles by means of an On-Board Unit. The On-Board Unit will collect probe data periodically, such as position and speed, but it will also generate event messages based on the data collected from sensors in the car. A GPS receiver is used for positioning. Another data source is anonymous mobile phone signaling data gathered from a cellular network to determine traffic parameters for certain road sections. In addition,
broadcasted TMC-RDS messages are used as data source. Urgent events generated in the Cooperative Cars equipped vehicle are forwarded to a Reflector Service, which immediately redistributes them back to all other Cooperative Cars vehicles in close vicinity. A Geocast Service broadcasts the events into a certain area, depending on the kind of event.

All events, periodically created probe data, mobile phone data, and TMC messages are sent to a database server, which cleans, transforms, and integrates the data using a Data Stream Management System. Data Stream Management Systems (DSMS) are well suited to cope with large amounts of data streaming in from sensors and other data sources and allow for an efficient continuous querying and processing of the data. The data is aggregated and stored in a historical database. This historical data can later be used to create and enhance quality of traffic information, such as produced in road traffic estimation and forecasting. A crucial component of the architecture is the Event Validator which allows for a prioritization and verification of messages. The Event Validator consists of several processing levels. While the first level just forwards urgent messages to the reflector service, the other levels verify the events by gradually using more information from the data streaming in, i.e., use data from a broader time interval. On each level of the validator, a quality measure is calculated, which expresses the level of reliability of the event potentially identified. Finally, the messages are distributed to the Cooperative Cars vehicles and a Web Service that can be used by external instances.
DATA FUSION

In the CoCar architecture different kinds of data sources are utilized to create valuable traffic information. The data streaming in from the different sources are integrated on-the-fly by using a Data Stream Management System. In the following, the utilized data sources and the envisioned fusion process of the architecture is detailed.

Data sources

The CoCar vehicles create messages both, regularly (probe data) and on demand (event messages). These messages are sent to the Reflector Traffic Information (RTI) Server, which streams them as JSON objects to consuming parties. The messages contain information that is common to all message types, such as the current position or the acceleration and also information which is specific to each type of event. For probe data the TPDP (Traffic Probe Data Protocol) and for event messages the FTAP (Fast Traffic Alert Protocol) is used. TPDP and FTAP are proprietary protocols defined in the CoCar project [23] while TPEG provides a standardized protocol framework to distribute traffic related information. An example
for an event type is *Emergency Braking*, which indicates that the current car is braking very hard (having an abnormal high deceleration). Other event types may be messages announcing traffic jams or an accident.

The second important data source is traffic data acquired and calculated from anonymous mobile phone probes. The idea is to derive traffic parameters for a road segment from vehicles carrying a mobile phone and travelling on that segment. This data source poses several issues regarding the accuracy of the measured traffic data. Firstly, a mobile phone has to be located on the road network. There exist multiple positioning techniques in a cellular network that are capable of locating a mobile phone. These positioning techniques vary, amongst others, in accuracy, costs, and supported cellular network [2]. Secondly, to identify on which road segment the vehicle carrying the mobile is travelling, a technique called *Map Matching* has to be applied. Depending on the used technique, such as simple map matching or topology map matching, the accuracy of the determined road segments may vary [9]. Other problems in deriving valuable traffic information from mobile phone probes are introduced by the occupancy of a vehicle, e.g., a bus could carry 20 travelers and a bus driver, all of them carrying a cell phone. Furthermore, pedestrians, bicycles, or trains can also introduce inaccuracy in measuring traffic parameters [7]. To better understand and oppose these problems, we develop a quality model, which will determine the varying errors introduced by the different technologies used. The simulation test bed we use is based on the work of [7] and is depicted in Figure 2.

The main idea of this simulation is to identify those characteristics of a WLT-based monitoring system that guarantee “good” estimations. The data from the simulated traffic scenario is degraded based on several parameters (e.g., quality of the positioning, degree of vehicles equipped with mobile phones). The degraded data is then used to estimate the mean speed on the roads. The quality of the estimated results is computed by comparing the estimated results with the true data of the traffic simulation. By simulating various scenarios and analyzing the generated quality, we can estimate the quality of WLT-based monitoring systems based on their characteristics (e.g., number of mobile phones in a certain area, accuracy of the positioning).
Thirdly, data from RDS-TMC, i.e., messages in the ALERT-C format [10] will be utilized to enhance and assess the quality of the traffic state estimation.

Lastly, historical data that were archived during the continuous operation of the system is used as a reference basis to enable and enhance quality of traffic state estimation and for traffic forecasting. One common way of historicizing traffic data is the time variation curve classification, where basic and detailed classification schemas can be distinguished. We will use a detailed classification described in [14] where known events, such as weekdays, public holidays or road works, as well as unpredictable events, such as bad weather conditions or accidents, constitute different classes. The current time variation curves can then be compared with the archived reference time variation curves. This method requires that data is gathered over a longer period of time. The data for the time variation curves is aggregated in fixed time intervals, e.g., 15 minutes.

**Data cleaning and fusion using a Data Stream Management System**

A *Data Stream Management System* (DSMS) is a data processing software handling continuous data streams such as sensor data. In contrast to common Database Management Systems, a DSMS encounters each data record only once, or in *one pass*. Queries defined in a DSMS query the data streams continuously and are usually defined over a window of records. This window may contain a constant number of records or is defined over a constant time interval, such as querying the last five minutes of the stream (indicated in Figure 3). The special features of a DSMS query, such as window width and the sliding step of the window, are often incorporated in the query language. SQL with extensions for data stream operators is
often used as a query language. The usual semantics of SQL needs to be adapted since blocking operators, like joins cannot be used on an indefinite set of records. Examples for Data Stream Management Systems are STREAM developed by the Stanford University [5], GSN (Global Sensor Network) created at the École Polytechnique Fédérale de Lausanne (EPFL) [6] and Aurora [6] formally known as Borealis. The interested reader can find more in-depth information about DSMS in [11,12,13].

In the CoCar data processing and fusion architecture we plan to use the GSN system as DSMS, as it is easily extendable for new data sources, it is distributable and it offers a lot of built-in functionality that can be reused for the CoCar architecture. To ensure a homogenous system, as much functionality as possible to manipulate a data stream should be integrated in the data stream flow. At first, a wrapper has to receive the data and make a first check, if the data is not corrupted. The wrapper can already perform some cleaning of the data, if necessary. Optionally, further cleaning and transformation steps are executed. Afterwards, the cleansed data proceeds to aggregation nodes, where it is aggregated for archiving and integrated with data from the other sources. The data is aggregated on the basis of time stamps and road segments. It is also possible to derive trends and summaries in data streams by using synopses. There exist different ways to create synopses from data streams, such as wavelet- and histogram-based techniques [13].

**USING PRIORITY AND QUALITY MEASURES**

From the point in time a data value enters the system it will be rated and prioritized. For each criterion in the system, a quality value and a weight for that criterion is defined. Furthermore, a priority, determining the urgency of the information is provided. At each stage in the data stream flow the quality measure is re-computed. In the first stage, when the record enters the DSMS, a quality measure rating the reliability of the data source is introduced.

Also in the cleaning and aggregation step a quality rating is performed. For data from floating vehicles, such as the messages created by CoCar vehicles or by Floating Phone Data, it has to be considered how many probes have been available for the aggregated value. Few probes make the aggregated values less reliable than values aggregated from many probes [15]. The aggregated data can be archived as historical data and used for the traffic state estimation and forecasting. Furthermore, event messages are passed on to the Event Validator component which rates the messages flowing in. Each message already has a reliability value and a priority when it reaches the Event Validator. The Event Validator component consists of several levels which are supposed to finalize the rating and uses the prioritization to
decide, how many evaluation levels the message may go through before it is forwarded to target vehicles. If the event is very urgent, such as an emergency braking, then the message is sent at once to vehicles in close vicinity of the message’s originating vehicle. In any case, the messages are subsequently streamed to the next level. For each level a fixed past time interval is specified. For the messages in this time interval the Event Validator verifies if the same event has also been reported by other vehicles. The more vehicles report the same event the higher the reliability value gets.

The quality measure will be based on the framework for data quality measurement of the U.S. Department of Transportation [19], which recommends the determination of six fundamental measures for each traffic data source: accuracy, completeness, validity, timeliness, coverage, and accessibility.

**RELATED WORK**

The CVIS (Cooperative Vehicle-Infrastructure Systems) is a Participant-to-X peer-to-peer architecture. Participants in the architecture can be peasants, vehicle drivers or other mobile users. Other components in the architecture comprise Traffic Management Centers, or roadside infrastructure, such as traffic control systems and signals [16]. The communication is realized over an IPv6 network infrastructure, which is very costly as the network infrastructure has to be provided and every participating component (such as traffic lights) has to be able to connect to the network. Furthermore, the architecture also acquires probe data from moving vehicles and integrates it with data from roadside sensors and detectors, which are themselves costly in deployment and maintenance. The data is integrated inside a subproject component called COMO (Cooperative Monitoring). It integrates and stores aggregated data with geospatial information and also provides some traffic state estimation facility.

The CO-Operative SystEms for Intelligent Road Safety (COOPERS) project funded by the EU aims at improving road safety by means of new traffic control applications and vehicle-to-infrastructure communication. In the COOPERS project different communication techniques will be investigated, such as DAB (Digital Audio Broadcasting), CALM (Communication Access for Land Mobiles) or IP networks [18]. The COOPERS architecture utilizes different data sources, such as static road sensors, Floating Car Data or environmental data, depending on the targeted application. Also comprehensive data quality assurance and reliability plans have been elaborated based on [19]. For each combination of measure, data source and COOPERS service an acceptance value has been defined.
The project Network on Wheels (NoW), successor of the FleetNet project and funded by the German Ministry of Education and Research proposes a Car-To-X architecture which is based on Wireless LAN and GPS. X means, that also car-to-car as well as communication between vehicles and road side infrastructure is included. The applications targeted in the NoW project range from safety enhancement, such as emergency vehicle or hazard warnings, to entertainment solutions which are integrated in the vehicles and the roadside units [20].

The CarTel system [21] is a distributed mobile sensor computing architecture incorporating a Data Stream Management System called ICEDB. Continuous queries are issued on mobile nodes (vehicles) that determine the kind and rate of sensor data to be acquired. For each sensor an adapter is used, which describes the schema, priority and acquisition mode, such as pull or push, by the ICEDB server. The local results are buffered and send or retrieved at the next possible point in time. Also CarTel applications can issue queries centrally over an ICEDB interface. The ICEDB system defines its own SQL-based continuous query language. In the CarTel system data can be prioritized locally (based on data on the nodes) or globally (based on a global view of all data in the system at the ICEDB server). Both methods change the prioritization dynamically based on the former data. As communication medium WiFi technology is used.

CONCLUSION AND FUTURE WORK

We propose a novel quality- and priority-based architecture for road traffic data fusion and processing in the context of the Cooperative Cars (CoCar) project. Data from mobile sources, such as mobile probe data from cooperative vehicles, anonymous Floating Phone Data and TMC messages are integrated to estimate and forecast traffic conditions also utilizing aggregated historical data. Furthermore, event messages detected by the CoCar clients are prioritized according to the event they represent and are assessed in a multi-level validation process. Based on different factors, such as the type of data source, the “cleanliness” of data, or the amount of information used to affirm an event, a quality measure rates the accuracy of the information included in the messages. The envisioned architecture is advantageous in comparison to other architectures because it enables cost-effective, fast, and accurate processing of traffic data and determination of current and prospective traffic states. Based on the concepts and principles presented in this paper, we are developing a prototype of the architecture, which will be tested and evaluated in the context of the Cooperative Cars project. We will conduct studies both in simulation and real-world setups to show the effectiveness of the architecture. In the simulation, to be able to test all features, we have to take into account Floating Car Data, event messages, as well as Floating Phone Data. Thus, besides simulating equipped
Cooperative Cars, cellular networks have to be simulated. As the equipment rate of Cooperative Cars rises, we will also carry out case studies at deployment test sites.

ACKNOWLEDGMENTS

This work has been partly funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) under the grant 01BU0690. The authors are responsible for the content of this publication.

REFERENCES


protocol for Radio Data System -- Traffic Message Channel (RDS-TMC) using ALERT-C


