Mining the Traffic Cloud: Data Analysis and Optimization Strategies for Cloud-Based Cooperative Mobility Management

Jelena Fiosina, Maksims Fiosins and Jörg P. Müller *

Abstract Future Internet (FI) technologies can considerably enhance the effectiveness and user friendliness of present cooperative mobility management systems (CMMS), providing considerable economical and social impact. Real-world application scenarios are needed to derive requirements for software architecture and smart functionalities of future-generation CMMS in the context of the Internet of Things (IoT) and cloud technologies. The deployment of IoT technologies can provide future CMMS with huge volumes of real-time data that need to be aggregated, communicated, analysed, and interpreted. In this study, we contend that future service- and cloud-based CMMS can largely benefit from sophisticated data processing capabilities. Therefore, new distributed data mining and optimization techniques need to be developed and applied to support decision-making capabilities of future CMMS. This study presents real-world scenarios of future CMMS applications, and demonstrates the need for next-generation data analysis and optimization strategies based on FI capabilities.

Key words: Cloud computing architecture, ambient intelligence, distributed data processing and mining, multi-agent systems, distributed decision-making

1 Introduction

Increasing traffic and frequent congestion on today's roads require innovative solutions for infrastructure and traffic management. As the components of traffic systems

---

Clausthal University of Technology,
Institute of Informatics,
Julius-Albert Str. 4, D-38678, Clausthal-Zellerfeld, Germany
e-mail: {Jelena.Fiosina,Maksims.Fiosins}@gmail.com, Joerg.Mueller@tu-clausthal.de

* The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. PIEF-GA-2010-274881.

become more autonomous and smarter (e.g. vehicles and infrastructure components are now equipped with communication capabilities), there is an increasing need for cooperation among intelligent transportation systems (ITS) for traffic management and environmental monitoring in order to improve traffic management strategies. Further, there is growing interest and increasing volume of investments to cooperative mobility management systems (CMMS). In these new-generation business management systems, the management of transportation networks is closely integrated with the business strategies and operational models of transport companies and individual customers, providing a considerable impact for companies in terms of business planning, service quality and adaption to customer needs as well as for individual users in terms of time and money saving, adaptive travel planning and support of social mutually beneficial behavior. Innovative cloud services can be created using the cloud capabilities of future Internet (FI) to access smart objects via the Internet of Things (IoT). This development can enable wide access to necessary information, because all of this data will be available in-the-cloud.

However, implementing a traffic cloud is far from easy. From an end user's point of view, the complexity of data and algorithms is hidden in the cloud. Users(ranging from traffic authorities to car drivers and automated components) expect to work with relatively simple applications on the Internet via mobile or embedded devices. These devices are fully connected and can (theoretically) use all the information available from all other users and system elements. This creates great opportunities for coordinated near-optimal management of the infrastructure (e.g. in terms of load balancing). However, there is a huge amount of available data with a short update rate. This creates a need to employ innovative data mining and corresponding decision-making algorithms (under the hood of the traffic cloud) to support CMMS applications in finding, collecting, aggregating, processing, and analysing information necessary for optimal decision-making user behavior strategies. Note that information here is virtually centralized by cloud technologies. However it is distributed, and (very often) created and managed in a decentralized fashion on the physical (fabric) layer. Thus, data mining and decision-making methods are required to find an optimal balance between decentralized information processing/decisions and costs of data transfer/decision coordination.

The contribution of this study is fourfold: First, we analyse related cloud-based architectures and CMMS scenarios. Second, we consider architectures for implementing the corresponding data analysis and optimization of mobility operations. Third, we discuss the employment of appropriate mathematical methods for three use-cases; fourth, we point out and discuss work directions and opportunities in the area of cloud-enabled CMMS.

The remainder of this paper is organized as follows. Section 2 reviews related work in the area of FI and cloud architecture for mobility application. In Section 3, we propose and analyse three application scenarios of CMMS and consider data analysis and optimization of participants' behaviour strategies in traffic systems. In Section 4, we present a cloud-based architecture for mobility networks based on the previously presented scenarios. Section 5 concludes and discusses future research opportunities.

2 Related Work

A strong worldwide interest in opportunities in transportation and mobility field has spurred the need for further analysing these FI opportunities. In Europe, FI and IoT research has been a priority direction for the 7th European Framework Programme (FP7) and will continue to do so for the upcoming Horizon 2020 Programme (e.g. the objectives 'A reliable, smart and secure IoT for Smart Cities' or 'Integrated personal mobility for smart cities' in FP7 or 'Substantial improvements in the mobility of people and freight' in Horizon 2020). These research questions are motivated and co-funded by private companies and municipalities from the areas of transport, logistics, communication and traffic management (e.g. the FP7 project Instant Mobility [**?**]). These stakeholders understand the possible enhancements to existing systems that new technologies can provide to CMMS. Research in this area is still largely at the stage of formulation of scenarios and coordination protocols. The first cloud-based traffic network architectures have been proposed in [**?**], which employ ambient intelligence (AmI) [**?**] or IoT components [**?**], [**?**].

An architecture of AmI-enabled CMMS is proposed in [**?**]. It supports virtual control strategists and management policy makers in decision-making and is modelled using the metaphor of autonomous agents. AmI is defined as the ability of the environment to sense, adapt, and respond to actions of persons and objects that inhabit its vicinity. Moreover, the multiagent system (MAS) paradigm makes AmI environments act autonomously and socially, featuring collaboration, cooperation, and even competitive abilities.

Cloud computing systems are oriented towards a high level of interaction with their users, real-time execution of a large number of applications, and dynamic provisioning of on-demand services. In this study, we consider the layered architecture of cloud-based computing systems presented in [**?**]. It supports a class of specialized distributed systems that is characterized by a high level of scalability, service encapsulation, dynamic configuration, and delivery on demand. The architecture includes the following layers:

The **fabric layer** includes all computing, storage, data, and network resources available in the cloud. The resources are accessible through the resource services, are used for cloud computations, management, and as testbeds. The **unified source layer** provides infrastructure-as-a-service by defining unified access to the raw computational resources of the fabric layer using a virtual machine. The **platform layer** provides platform-as-a-service, including a collection of specialized tools, middleware, and services on the top of unified resources to create a deployment platform (e.g. scheduling create service and artificial testbeds). The **application layer** contains all applications that are run in the cloud. Application execution in the cloud is distributed: applications can be partly executed on the client, partly in the cloud.

The application of cloud-based architectures for ITS is demonstrated in [**?**]. In order to provide an acceptable level of service, a cloud-based ITS consists of two main components: an *application component*, which provides dynamic services and runs all the cloud applications; and a *digital (simulated) traffic network* component, which performs constant information collection and processing in order to provide in-time data. A cloud-based ITS adapts its decisions by using available information and by interacting with human as well as automated traffic participants.

We apply our experience in implementing data processing, mining [**?**], [**?**], and decision-making methods [**?**], [**?**] for existing transportation problems. Next, we discuss the key aspects of methods in future-generation CMMS.

3 Traffic Cloud Scenarios and Related Data Analysis and Optimization Problems

We propose three cloud-based ITS application scenarios: 1) **A cooperative intersection control**, which optimizes vehicle flows in traffic networks by regulating the intersection controllers. 2) **A personal travel companion**, which provides dynamic planning and monitoring of multimodal journeys to travelers, surface vehicle drivers, and transports operators. 3) **A logistics services companion**, which provides benefits to clients and stakeholders involved in, affected by, or dependent on the transportation of goods in urban environments. We demonstrate the most important stages of data processing and optimization in order to derive requirements for a general architecture described in the next section.

3.1 Virtualized cooperative intersection control

This scenario uses adaptive, semi-distributed traffic management strategies hosted in the cloud for the regulation of intersection controllers, and creates ad-hoc networks in the cloud between clusters of vehicles and the traffic management infrastructure. It recommends the optimal speed to drivers to keep the traffic flow smooth, and assists adapting traffic controllers (e.g. traffic lights, signs) based on the real-time traffic situation. This service uses real-time traffic information and a route-data collection service to formulate strategies for the optimization of network operation.

Stage 1: Processing the following data streams (historical and real-time): 1) floating-car data (speeds, positions, etc.); 2) sensor data from the infrastructure (loops, traffic lights, etc.); 3) information about routes and actual locations of collective transport (public transport, taxi, shared cars, etc.) 4) data from distribution vehicles (logistic transport); 5) weather conditions; 6) accidents, car breakdowns, road-works; 7) organizational activities (sport events, conferences, etc.)

Stage 2: Creating ad-hoc networks, which are virtual abstract networks for solving specific problems (intersection and regional traffic models, green wave models, public transport priority, jam avoidance, etc.). Estimating network parameters (traffic flux, density, and speed, travel time estimation, etc.).

Stage 3: Developing static strategies of intersection control and cooperation based on historical information, previous experience, and data models from the previous stage (traffic light signal plan optimization; signal plans for expected events (such as increase of flows, changed weather conditions, organizational activities); cooperation plans of clusters of vehicles, etc.).

Stage 4: Combining dynamic real-time information with static strategies in order to receive up-to-date controlling decisions (correction of signal plans according to current conditions, cooperation of signal controllers to resolve problems such as jams, accidents, etc.)

3.2 Dynamic multi-modal journey planning

The purpose of this use case is to help travellers plan and adjust a multi-modal, door-to-door journey in real-time. It provides improved (i.e., quicker, more comfortable, cheaper, and greener) mobility to daily commuters and other travellers by identifying optimal transportation means and a strong real-time orientation. This planning proposal for a multi-modal journey takes into account the current means of transportation, the traveller's context and preferences, city traffic rules, and the current requirements and constraints. The journey plan needs to obtain an overall indication of the trip duration as well as accommodate early reservation of resources (train or plane ticket).

Stage 1: Processing of the following data streams (historical and dynamic) in addition to the previous application: 1) floating passenger data; 2) travellers' preferences; 3) timetables and availability of collective transport (tickets, shared cars availability, etc.); 4) changes in time-tables.

Stage 2: Creation of ad-hoc networks (transit stations, public transport coordination, passenger choice of transport, etc.) and estimation of network parameters (travel time for different transport modes depending on various factors, waiting times, passenger arrival at stops, price models, etc.).

Stage 3: Multi-modal route pre-planning based on historical data and estimated network parameters for expected conditions (pre-planning for popular routes, pre-planning for pre-booked routes, pre-planning for expected events) as well as optimal time-table calculation for public transport based on the expected conditions.

Stage 4: Dynamic update of pre-planned routes for the actual multi-modal journey (actual travel-time estimation, re-planning in the case of delays in previous trips in the multi-modal chain, re-planning for additional travel possibilities, or cancelling a part of the multi-modal journey), as well as dynamic update of public transport time-tables (on-demand changes, co-ordination of different transport means).

3.3 Itinerary booking and real-time optimized route navigation

This use case helps a logistics provider (1) guarantee quick (especially on-time) deliveries at a low cost based on up-to-date information and (2) maximize the efficiency of each vehicle and the fleet. It is fundamental to optimize the movements of the logistics vehicles, to help them avoid traffic jams and take the shortest routes when possible.

Stage 1: Processing of the following data streams (historical and dynamic) in addition to the first application: 1) order data (transportation demand); 2) available logistic vehicles (possible load, speed, etc.); 3) timetables (if necessary) and actual positions of the vehicles; 4) client data (drop-off preferences, actual location, etc.).

Stage 2: Creation of ad-hoc networks (delivery models, logistic provider-client interaction models, etc.), and estimation of the network parameters (travel times for different route segments, delay probability, drop-off process time distribution, probability of accidents, probability of problems with vehicles, etc.).

Stage 3: Pre-planning of the delivery process (preliminary good distribution by vehicles, preliminary order of clients for each vehicle, preliminary route for each vehicle, preliminary time window for each client, etc.). Note that the itineraries of large logistic operators can be used to provide better predictions of the traffic situation using virtualized cooperative intersection intelligence application as well as by applying priority rules for logistic vehicles during booking.

Stage 4: Dynamic update of pre-planned delivery routes depending on up-to-date information (re-planning of routes depending on current traffic situation, re-planning in the case of accidents or traffic jams, re-planning in the case of vehicle problems, estimation of actual delivery time, etc.). Cooperation between logistic vehicles (exchange or orders, adoption of other vehicle's orders in the case of problems, etc.). Dynamic agreement with clients (agreement about drop-off place depending on current position of the vehicle and client, agreement about change of drop-off time, reaction to the new/changed customer requests, etc.).

4 Reference Architecture for Traffic Cloud Data Mining and Strategy Optimization

The applications mentioned in the previous section are data-intensive. Services provided through the cloud require large amounts of data to be processed, aggregated, and analysed. Then, the processed data is used for calculating optimal strategies for traffic participants. Now we generalize the stages of data processing and network optimization from the scenarios discussed in the previous section. We propose a reference architecture for traffic cloud data mining and optimization of strategies (TCDMOS), which is based on [**?**], but we focus on data processing and decisions. TCDMOS is illustrated in Fig. 1. It includes the following stages of data processing and network optimization:

Stage 1: *Mining data from the IoT and its pre-processing*. All the participants of the cloud-based system have virtual representations as active IoT components (agents). These virtual agents are associated with data (mostly real-time) and act as data sources for the cloud-based system. The cloud system locates and collects the necessary data from different agents, and provides usual data mining operations (changes and outliers are detected, preliminary aggregation and dimensionality re-

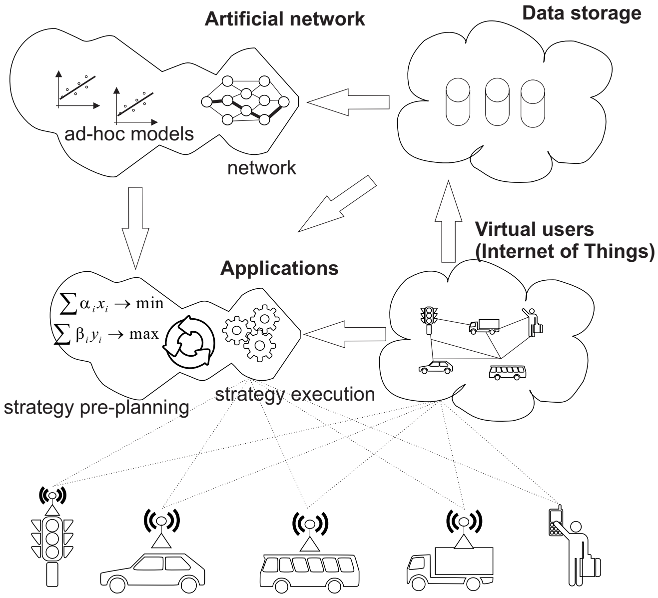

Fig. 1 TCDMOS Architecture: Traffic Cloud Data Mining and Optimization of Strategies

duction are performed). The collected data are stored as historical information in the cloud and are used later as input data for ad-hoc network models (Stage 2). Stream-based methods of semi-decentralized change-point detection, outlier detection, clustering and classification, and factor analysis occur regularly in this stage.

Stage 2: *Ad-hoc network models*. The application-specific digital networks of virtual traffic participants (e.g. regional, social) are created, and the corresponding data models are used in order to estimate the important characteristics and parameters of these networks using the information collected in Stage 1 and for strategy optimization at Stage 3. The future behaviour of traffic participants is forecasted as well. Semi-decentralized, flows forecasting (possibly with incomplete information) methods such as (multiple-response) regression models, Bayesian networks, time series, simulation, are also applied at this stage. Many pre-defined data models can run concurrently in the digital network. The corresponding data storages are located in the cloud and are semi-centralized, so the methods should take costs of different pieces of information into account.

Stage 3: *Static decisions and initial strategy optimization*. Cloud applications use pre-calculated results of the ad-hoc network models from Stage 2 and the available historical information (including private information) about the traffic network to perform their pre-planning tasks. Initial optimization of the strategies is resource-expensive, and can be partially pre-calculated in ad-hoc network models and then instantiated according to the application's goals and preferences. These models are also checked in the digital traffic network. This stage can require aggregation of different data models and existing strategies. Methods of self-learning stochastic (multi-criteria) optimization such as neural networks, decision trees, Markov decision processes, choice models, graph optimization algorithms are used.

Stage 4: *Dynamic decisions and strategy update*. The pre-planned tasks from Stage 3 are executed, and updates are made according to the dynamic real-time situation extracted from the virtual agents. The aggregation of the pre-planned data and strategies with the dynamic ones is the most important problem at this stage. An additional difficulty here is the requirement of fast real-time execution. (Automatic) cooperation between users in their decisions is possible; therefore, stream-based methods of data models and strategy updates such as reinforcement learning, Bayesian networks, dynamic decision trees, stream regression, and distributed constraint satisfaction/optimization can be applied.

5 Future Work and Conclusions

The main contribution of this study is a reference architecture for traffic cloud data mining and optimization of strategies (TCDMOS) and related data processing and network optimization methods. We envisage this as an important step towards making FI and cloud technologies usable for next-generation CMMS. TCDMOS requirements were elicited from traffic scenarios, which reflect needs and impact of CMMS for business and society, and the corresponding problems that should be solved for effective cloud system operation were illustrated. Future work will be devoted to elaborating the architecture, developing novel algorithms, and integrating and validating them in state-of-the-art cloud computing frameworks.

References

1. 7-th european framework programme project, instant mobility: Multimodality for people and goods in urban area, cp 284806. URL http://instant-mobility.com/
2. Fiosina, J.: Decentralised regression model for intelligent forecasting in multi-agent traffic networks. In: S.e.a. Omatu (ed.) AISC - 9th Int. Conf. on Distributed Computing and Artificial Intelligence (DCAI'12), vol. 151, pp. 255–263. Springer (2012)
3. Fiosina, J., Fiosins, M.: Distributed cooperative kernel-based forecasting in decentralized multi-agent systems for urban traffic networks. In: Proc. of Ubiquitous Data Mining (UDM) Workshop of ECAI 2012, pp. 3–7. Montpellier, France (2012)
4. Fiosins, M., Fiosina, J., Müller, J.P.: Change point analysis for intelligent agents in city traffic. In: Agents and Data Mining Interaction, *LNCS*, vol. 7103. Springer (2012)
5. Fiosins, M., Fiosina, J., Müller, J.P., Görmer, J.: Reconciling strategic and tactical decision making in agent-oriented simulation of vehicles in urban traffic. In: Proc. of 4th International ICST Conference on Simulation Tools and Techniques (SimuTools'2011) (2011)
6. Foster, I.: Cloud computing and grid computing 360-degree compared. In: Proc. of the Grid Computing Environments Workshop, pp. 1–10 (2008)
7. Li, Z., Chen, C., Wang, K.: Cloud computing for agent-based urban transportation systems. IEEE Intelligent Systems **26**(1), 73–79 (2011)
8. Passos, L., Rossetti, R., Oliveira, E.: Ambient-centred intelligent traffic control and management. In: Proc. of the 13th Int. IEEE Annual Conf. on ITS, pp. 224–229 (2010)