

COMMUNITY ACTIVITY: WORKSHOP REPORT

Towards a computational transportation science

Summary: This report of a community activity, a Dagstuhl Seminar earlier in 2010, postulates the need for a *computational transportation science*, as the science behind intelligent transportation systems. In addition to the argument for establishing a discipline, we present a first research agenda for computational transportation science.

1 An emerging discipline

In the near future, vehicles, travelers, and the transportation infrastructure will collectively have millions of sensors that can communicate with each other. This environment will enable numerous new applications and dramatic improvements in the performance of existing applications. Due to their distributed and mobile nature, future transportation systems may become the ultimate testbed for a ubiquitous (i.e., embedded, highly-distributed) and sensor-laden computing environment of unprecedented scale.

This field is currently subsumed by intelligent transportation systems, or ITS. However, the question arises whether behind intelligent transportation *systems* we also need a *science* (for a similar discussion see [2]). The paradigm shifts witnessed in technical possibilities—for example, from centralized to distributed or decentralized computing, from carefully managed authoritative data to massive real-time data streams of unknown quality—may require new scientific foundations. More and more aspects of transportation science require sophisticated computational methods to deal with the complexity of dynamic environments. We argue that a better interface between transportation science and computer and information science is needed. The communication and exchange between these scientific communities would improve, but there are also shared common themes and long-term research questions. Around these core research themes we define a new discipline: *computational transportation science*.

Computational transportation science (CTS) concerns the study of transportation systems where people interact with information systems (e.g., interfaces for driver assistance, or integrated transport information); where systems monitor and interpret traffic (e.g., mining for activity patterns, or crowd-sourcing to monitor events); or where systems manage the traffic (e.g., control of traffic flow at traffic lights, or toll management). CTS inherits from computer science the aspects of distributed and decentralized computing and spatiotemporal information processing, and from transportation science the aspects of transportation

control and management. The discipline goes beyond vehicular technology, and addresses pedestrian systems on handheld devices, non-real-time issues such as data mining, as well as data management issues above the networking layer. CTS studies how to improve the safety, mobility, efficiency, and sustainability of the transport system by taking advantage of information technologies and ubiquitous computing. In particular it needs scholars and practitioners that maintain a body of knowledge and push forward an agenda that is deeply rooted in both established disciplines. We are also the first to admit that drawing the lines between the established and the emerging discipline is to some extent arbitrary. The intention of claiming an emerging discipline is by no means exclusive or divisive; rather it is to draw together work that is otherwise disconnected, and to foster research in this area through recognition. In all the examples above, and in the research agenda below, there exists research that is already underway.

In this way, CTS becomes the science behind ITS. Academic ITS communities, such as in the IEEE ITS Society (founded 2005), already interpret the “S” as science, not systems (otherwise they would not pass scientific peer review). However, the scientific discipline behind ITS cannot be named intelligent transportation science in analogy to ITS—there is no such thing as an “intelligent science,” and then there is also an established discourse in artificial intelligence whether machines can be intelligent. So “computational transportation science” seems to say it all.

2 History

Computational transportation science has made its first steps of consolidation. A PhD program on the subject, funded by the National Science Foundation, was established at the University of Illinois at Chicago in 2006. Two international workshops on CTS were held (2008 in conjunction with the *5th Annual International Conference on Mobile and Ubiquitous Systems*, and 2009 in conjunction with the *17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*). A third workshop was held in conjunction with the *18th ACM SIGSPATIAL International Conference* in 2010. With the first workshop appeared a preliminary publication exploring a research agenda in this area [1]. Then a Dagstuhl Seminar on computational transportation science was held in 21–26 March 2010 to characterize the discipline and identify its research agenda. The seminar was attended by 25 invited researchers from USA, Australia, Germany, Belgium, and Switzerland, with nationalities also from China, India, Greece, and former Yugoslavia. This report presents the highlights of this Dagstuhl Seminar. Major steps at the seminar have been:

- agreeing on a collaborative definition of CTS, vision of CTS, and core research agenda for CTS;
- setting up a Wikipedia entry for the definition and vision¹;
- setting up a webpage as a bulletin board for the growing community²;
- preparing and holding the third international workshop on CTS in conjunction with ACM SIGSPATIAL GIS 2010;
- engaging with funding bodies promoting CTS as a discipline (outreach);
- establishing collaboration by developing some larger joint research project proposals;
- publishing the (first) core research agenda via this report.

¹http://en.wikipedia.org/wiki/Computational_transportation_science, last accessed 5 June 2010

²<http://www.ctscience.org/>, last accessed 5 June 2010

3 Core research agenda of CTS

A discipline is, among other properties, characterized by a common core research agenda. Computational transportation science must have some underlying long-term fundamental research problems to distinguish it from its application area, ITS. ITS, in comparison, has defined several research and development agendas that are typically short-term, such as the US DOT's strategic ITS plan [7].

The following research agenda is the outcome of the discussion of the Dagstuhl Seminar on computational transportation science. As such it is preliminary and biased by the composition of the participants. Nevertheless it demonstrates vision and need for this discipline. This preliminary agenda is structured into five sections: applications, knowledge discovery, decentralized computing, social computing, and societal issues.

3.1 Applications

As the name "computational transportation science" indicates, important aspects are computational and algorithmic aspects in CTS. The challenges lie in the diversity of sensors and thus data gathered in different spatial, temporal, and thematic resolution. The high volume demands for adequate information reduction for processing. One way to solve it is to exploit the principle of locality, i.e., the fact that information is mainly relevant locally and thus can also be processed locally and need not be communicated and processed on a central server. This leads to concepts of decentralized and distributed processing.

The applications described below rely not only on the fact that travelers are provided with information; as travelers are equipped with sensors capable of acquiring information of the local environment, travelers also act as data providers. This leads to a highly dynamic map of the environment that can be exploited in numerous ways. On the one hand, the technology provides real-time data and thus can be used for dynamic traffic assignment; on the other hand, the technology can also enhance the perception range of individuals and allow them to "look around the corner," or to "look through the cars in front of them." An additional important benefit is the possibility to augment the environmental information with virtual information about the infrastructure. In this way, virtual traffic lights or virtual lane assignments can be realized to allow for a flexible traffic management.

Not only the data can be shared, but also the transportation resources can be shared. This is already the case for the road network and for public transportation. However, sharing can also be envisaged for other vehicles, like private cars.

The applications are driven by different factors:

- The ever increasing traffic demand leading to congestion, with dramatic effects on public safety, the environment, and the economy due to time spent in traffic jams.
- Real infrastructure is expensive and laborious to maintain; furthermore, it is ageing and has to be replaced by modern concepts and systems.
- Cars and travelers are increasingly equipped with sensors which can—among other things—capture information about themselves and about the local environment. This rich data source can be exploited.

In the following, some future applications are described:

1. Shared transportation resources: If all traffic modes are considered (including private traffic), a better exploitation of the resources is achieved, with several benefits for the

- users (reduced prices), the infrastructure (less congestion), and as a consequence also the environment (less pollution).
2. Collaborative travelling: Collaboration can be used for platooning: the virtual coupling of vehicles to form larger units like virtual trains. These structures can get priorities, e.g., when crossing junctions. Within a platoon, autonomous driving is possible. Further, there are opportunities for more adaptive traffic management depending on the current traffic situation (e.g., intersection negotiation and intelligent traffic lights).
 3. Physical infrastructure is replaced by virtual infrastructure: Virtual infrastructure can offer several advantages over ageing and expensive to maintain physical infrastructure. For example, virtual lanes can compensate for different traffic volume during the day/week. Virtual traffic lights and virtual signs may be possible, as well as transient and ad-hoc warnings, like construction sites, aquaplaning, or slippery roads.
 4. Driver assistance: Drivers can be warned of risks in their local environment or when risking to leave their lane. Furthermore, drivers' visibility range can be expanded by providing up-to-date information from areas that are currently invisible.
 5. Evacuation planning: Highly temporal information is provided to support and calibrate simulations, with the objective of emergency preparedness.
 6. Autonomous driving: As a long-term goal, highly dynamic maps of the environment have the potential to support autonomous driving.
 7. Dynamic road pricing: Knowledge about the current usage of roads can be used to manage traffic, e.g., by reducing prices for collaboratively used cars or platoons.
 8. Smart grid, electric cars: Sharing resources opens the way to extend the flexibility of using and sharing electric cars, e.g., by dynamic planning of the electric grid resources, and of routes by considering charging facilities.
 9. Road and traffic planning: Road and traffic planning can be greatly enhanced by precise, high resolution travel information, which leads to adaptive traffic systems. For example, the road visibility, precipitation, and pavement condition information can be provided at high spatial resolution.

In general, the major benefits and expected properties are robustness (due to high redundancy of information), resilience (ability to recover after failure), reliability, and timeliness, which is relevant both for offline and online applications described above.

3.2 Knowledge discovery, filtering, and visualization

In order to be efficient, safe, and environmentally friendly a traveler must be cognizant of their inherently dynamic surroundings both through their own sensing systems and by communicating with other travelers and systems. At present travelers gain most of their situational awareness from their innate sensors (eyes, ears, etc.) perhaps augmented by delayed reports from the radio or the web.

Thus it is important to discover in a timely fashion additional information that can augment the innate sensors. Consider for example the query: What are the expected traffic conditions on I298 at Ontario over the next hour? This query can be answered by a server that stores historical information; but additional information may be available on the web, e.g., the weather and special events such as a ball game that starts at the time. In this case it is not even clear what data and web sites are relevant to the query.

It will not be long, however, before the traveler will be inundated with real-time information coming from all distance scales over soon-to-be ubiquitous always-on wireless

networks. Prioritization of messages will be critical. Hence, knowledge discovery, filtering, and visualization form important research challenges. There is a need to devise mechanisms that can assist in making sense of the huge volumes of heterogeneous and distributed data in particular decision making contexts. Some challenges are:

1. For the car driver, the vehicle itself will not only be aware of the vehicles around it, due to a plethora of on-board sensors (such as radar and computer vision systems), but also of their intentions through constant DSRC (dedicated short-range communications) exchanges. Road signs will become redundant as the data is sent directly to the vehicle. Speed limits will be mandatorily controlled to increase the safety and efficiency of roads. Such a detailed and consolidated picture of the local environment around the vehicle has the potential to reduce the number and severity of collisions and so increase traveler safety. Others, such as pedestrians and cyclists, will be similarly equipped, communicating constantly with nearby vehicles and travelers. Thus the safety of even the most vulnerable road users will be enhanced.
2. In large cities and on congested roads the data density will be vast. For individual travelers, and the devices and systems that are assisting them in making a journey, only a small fraction of the received data will be relevant (and even less will be useful) and some form of stream processing will be necessary just to prioritize the messages in order of immediacy and relevance let alone acting on their content. That is not to say that any of the data is useless. Indeed, it contains trends and anomalies that are useful for planning not just the next trip but also the transportation capacities required in the future. Extraction of these trends and anomalies must be automated and conveyed to the relevant user in an easily understandable form.
3. Since there is no guarantee that the data available to a traveler are of useable quality or even available when needed, filling the spatial and temporal data gap is a challenging issue. Is it meaningful to fill the gaps with data from yesterday or even a minute ago? Can statistical machine learning techniques such as support vector regression help? The answers are not clear and must depend on what the data is to be used for. After all a bus timetable is simply a prediction of often dubious reliability.
4. Visualization of the huge, multi-dimensional data sets generated will not be easy. Many users will have their own requirements and will want to construct queries and visualize the results. It is unlikely that the mobile device of an individual user will have the computational power or storage for such a task. Will cloud computing come to the rescue? Will peer-to-peer systems help with data storage and download? The physical presentation of the data is also an issue. An in-vehicle display must not be too obtrusive or interfere with the driver's ability to control the vehicle (at least until the vehicle is fully autonomous). Questions of relevance, urgency, and safety need to be addressed.

There are numerous unanswered questions raised in the paragraphs above. They are potentially solvable in isolation but all the possibilities will only be leveraged through unified study. CTS is aimed at this broad field.

3.3 Decentralized computing

While applications are concerned with what data is relevant to answer a particular query, decentralized computing is concerned with where this data resides and how to access this

data. First we will explain what these questions mean in the CTS environment, and then why these questions are different than the ones answered by traditional DBMSs.

Consider for example the query: What is the average speed of traffic a mile ahead of me? Sometimes the query can be posed to a central server, but sometimes a server with this information is unavailable (e.g., because the query pertains to a congested side street that is not instrumented with speed sensors), and the query needs to be answered by polling the vehicles ahead. However, the network identities of these nodes (the vehicles) are not known. Thus, for this query it is not known where the data resides, and how to get to it. The answer in this case may be to use short-range wireless communication such as WiFi or DSRC to disseminate the query to neighboring nodes transitively. In other words, the limited transmission range of the network is used to compensate for the lack of knowledge of node identity.

These questions are not addressed by traditional DBMSs. The data integration problem studied by the database research community assumes that the data is always available, but the integration part is the problematic part. In distributed databases it is assumed that there are directories that map data to network identities of computers that store the data. Such directories are appropriate for some of the data that pertains to the query, but certainly not to queries that are processed by polling other vehicles.

The research challenges in decentralized computing are:

1. Modeling and representation of highly heterogeneous data, including: the context dependency of data and the reconciliation in different contexts; the (semantic) interoperability; dynamic management of interoperability; and data warehousing.
2. Methodologies/tools to orchestrate and control the information flow in and across infrastructure networks, including: quality of data guarantees and control of local or partial knowledge, and quality measures for cross-layer execution of decentralized algorithms and data integration; satisfying resource constraints such as bandwidth, energy limits, or computational capabilities; aiming for global optimization in transportation from local clusters and local knowledge.
3. Novel information-theoretic formalisms and measures for complexity and efficiency, for example, data freshness/staleness, communication costs, and transitions among different levels of hierarchical decentralization.
4. Data reduction and aggregation, for example, in the context of streaming, storage, and management across hierarchical layers needs to be addressed, where the structure of the different layers can evolve or change over time.

3.4 Social computing

Social computing and information processing taps into the wisdom of crowds [6], and relies on the (ubiquitous) connectedness and communication ability of the members forming the society. Provided such infrastructure exists, cooperation in terms of computing and information processing becomes feasible and forms new research questions [4].

While social computing is making its way into many disciplines, it is obvious how real-time social interaction between mobile and stationary individuals (people, vehicles, goods, and infrastructure) can improve transportation. From a state where every individual is acting autonomously in isolation, or with minimal (visual) interaction with their environment, it is quite a paradigm shift to think of transportation as an interconnected, communicating, and cooperating complex system (e.g., [5]).

Such a paradigm shift brings up research questions in multiple dimensions:

1. Data collection. How can alternative data collection methods to the traditional survey-based collection assist in computational transportation science? This involves direct and indirect collection methods such as real-time sensing or accessing pools of shared data, and issues of data integration and exchange.
2. Data sharing. How can we ensure a sustainable data flow? Why do individuals share data, what are the incentives? What type of data will be captured and shared, and what type of data needs protection?
3. Data quality assurance. How can social network based data be calibrated, or errors identified? How can trust and reliability inform an error propagation process?
4. Real time situational awareness and decision support. How can data be transformed into information, and provided where and when needed for decision support?
5. Privacy aspects. How can privacy be protected when spatial information provides local information about the user? Which computational methods can be devised in order to blur the local information and at the same time still keep it useful? It should also be noted that with the ever decreasing cost of storage it may well be practical to maintain all the data forever. But who controls it? Should an arbitrary person or agency be able track an individual's movements? Legislation may mandate privacy but can't guarantee it. Also data privacy impacts usability. For example adding noise to the position of a vehicle would render the safety aspects of DSRC useless. Removing personal identifiers is one possible solution, but is it the best?

Ultimately it is the traveler who specifies the origin and destination of their trip. Transport must be traveler-centric and any system (or combination of systems) that ignores this simple fact will have a limited life. It is the socioeconomic outcomes that applications of CTS will generate that will really count in the long run.

Abstracting from social computing, other societal issues of computational transportation science come up with the involvement of individuals in the information processing chain. These individuals act and interact in a larger societal context, involving also government (transportation authorities), stakeholders, and transportation providers. These groups have different roles and responsibilities, but also different values and interests.

Research questions in this area concern the complex decision making processes, economic models including novel fare models (e.g., e-tolling, e-ticketing, ride sharing, virtual fencing), and also the demands of the community for privacy.

4 Outlook

A discipline is only as good as its academic community. If this paper finds your support or meets your interests you are cordially invited to participate and engage. The infrastructure set up so far is a beginning but requires your collaboration, be it the Wikipedia entry, the CTS webpage, or the CTS workshop series. These are all small seeds that—if they grow—can lead to conferences and journals on CTS, not only in the content but also in name. Finally, the community should shape its own academic programs or introduce core subjects on computational transportation science into the programs on transport engineering, electrical engineering, software engineering, and geographic information engineering. The spread demonstrates the interdisciplinarity of computational transportation science,

illustrates that engineering problems do not present themselves any longer wholly contained in one traditional discipline, and supports the fundamental concern that engineering disciplines have grown to be too narrow [3].

Acknowledgements

The authors are grateful to a number of institutions that supported this work. Dagstuhl Seminars are sponsored by Schloss Dagstuhl, an institute in the Leibniz Association, Germany. Work and travel was further funded by the Australian Research Council (DP0878119), National ICT Australia, and the National Science Foundation of the USA through awards DGE-0549489 and IIS-0957394. The authors would also like to thank all contributors at the Dagstuhl Seminar, especially Steve Liang for setting up the CTS webpage and Bo Xu for creating the CTS Wikipedia entry.

References

- [1] GEERS, D. G. Some research questions for computational transportation science. In *Mobiquitous '08: Proc. 5th Annual International Conference on Mobile and Ubiquitous Systems* (Brussels, 2008), ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), pp. 1–5. doi:10.4108/ICST.MOBIQUITOUS2008.3830.
- [2] GOODCHILD, M. F. Geographical information science. *International Journal of Geographical Information Systems* 6, 1 (1992), 31–45. doi:10.1080/02693799208901893.
- [3] NATIONAL ACADEMY OF ENGINEERING. The engineer of 2020—visions of engineering in the new century. Tech. rep., National Academy of Engineering, 2004.
- [4] PARAMESWARAN, M., AND WHINSTON, A. B. Research issues in social computing. *Journal of Association for Information Systems* 8, 6 (2007), 336–350.
- [5] SORENSEN, J. One in every crowd. *Traffic Technology International* 2 (2010), 50–56.
- [6] SUROWIECKI, J. *Wisdom of the crowds*. Doubleday, New York, 2005.
- [7] UNITED STATES DEPARTMENT OF TRANSPORTATION RESEARCH AND INNOVATIVE TECHNOLOGY ADMINISTRATION (RITAS). ITS strategic research plan 2010–2014. Tech. rep., United States Department of Transportation, 2009.

Stephan Winter
University of Melbourne, Australia
Monika Sester
Leibniz University Hannover, Germany
Ouri Wolfson
University of Illinois, Chicago, USA
Glenn Geers
National ICT Australia, Sydney, Australia